



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2024년10월22일

(11) 등록번호 10-2720785

(24) 등록일자 2024년10월18일

- (51) 국제특허분류(Int. Cl.)
C12Q 1/6827 (2018.01) **C12Q 1/6886** (2018.01)
G16B 30/00 (2019.01)
- (52) CPC특허분류
C12Q 1/6827 (2018.05)
C12Q 1/6886 (2022.01)
- (21) 출원번호 **10-2024-7011005(분할)**
- (22) 출원일자(국제) **2013년06월14일**
 심사청구일자 **2024년04월04일**
- (85) 번역문제출일자 **2024년04월02일**
- (65) 공개번호 **10-2024-0049635**
- (43) 공개일자 **2024년04월16일**
- (62) 원출원 특허 **10-2023-7012164**
 원출원일자(국제) **2013년06월14일**
 심사청구일자 **2023년04월19일**
- (86) 국제출원번호 **PCT/IB2013/054898**
- (87) 국제공개번호 **WO 2013/190441**
 국제공개일자 **2013년12월27일**
- (30) 우선권주장
 61/662,878 2012년06월21일 미국(US)
 (뒷면에 계속)
- (56) 선행기술조사문헌
 WO2012071621 A1

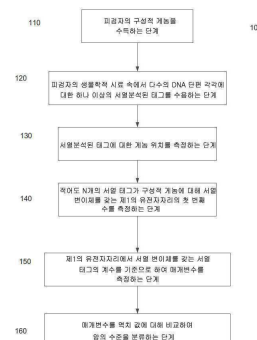
- (73) 특허권자
더 차이나이즈 유니버시티 오브 홍콩
 중국 뉴 테리토리즈 홍콩 새턴 더 차이나이즈 유니버시티 오브 홍콩 피 치우 빌딩 오피스 오브 리서치 앤 놀리지 트랜스퍼 서비스즈 룸 301
- (72) 발명자
로 유크 밍 데니스
 중국 홍콩 (에스에이알) 호만틴 코울룬 킹 탁 스트리트 7포스 플로어
- 찬 판 치**
 중국 홍콩 에스에이알 카울룬 메이 푸 선 쉐엔 브로드웨이 스트리트 블록 34 13/에프 플랫 에이
 (뒷면에 계속)
- (74) 대리인
김진희, 김태홍

전체 청구항 수 : 총 28 항

심사관 : 김정태

(54) 발명의 명칭 **암 검출을 위한 혈장 DNA의 돌연변이 분석****(57) 요약**

암을 스크리닝하거나 모니터링하는 피검자의 생물학적 시료(예를 들면, 혈장 또는 혈청)내 체세포 돌연변이의 빈도를 동일한 피검자의 구성적 DNA 내의 것과 비교할 수 있다. 매개변수는 이들 빈도로부터 기원할 수 있으며 암의 수준의 분류를 결정측정하는데 사용될 수 있다. 거짓 양성은 임의의 변이체 유전자좌가 적어도 규정된 수의 변이체 서열 관독물(태그)을 갖도록 요구함으로써 필터링되어 적어도 규정된 수를 갖기 위해 어떠한 변이체 유전자좌를 요구함으로써, 보다 정밀한 매개변수를 제공하여 필터링 제거할 수 있다할 수 있다. 상이한 변이체 유전자좌에 대한 상대적인 빈도를 분석하여 환자내 종양의 이중성의 수준을 결정측정할 수 있다.

대표도 - 도1

(52) CPC특허분류

G16B 30/00 (2019.02)

C12Q 2537/165 (2013.01)

C12Q 2600/156 (2013.01)

(72) 발명자

지앙 페이용

중국 홍콩 에스에이알 뉴 테러토리즈 새틴 시우 렉
유엔 로드 노스 62-66 쉥 람 코트 블록 비 퍼스트
플로어 플랫 7

치우 와이 켄 로사

중국 홍콩 뉴 테러토리즈 새틴 마 록 패스 52 더블
헤이븐 하우스 31

(30) 우선권주장

61/682,725 2012년08월13일 미국(US)

61/695,795 2012년08월31일 미국(US)

61/711,172 2012년10월08일 미국(US)

13/801,748 2013년03월13일 미국(US)

명세서

청구범위

청구항 1

피검자 내에서 암 또는 전암성 변화를 검출하는 방법으로서, 상기 방법은 하나 이상의 프로세서를 포함하는 컴퓨터 시스템에 의해 수행되고,

피검자의 생물학적 시료 중의 다수의 DNA 단편 각각에 대한 하나 이상의 서열 태그를 수취하는 단계로서, 생물학적 시료는 무세포 DNA를 포함하는 것인 단계;

다수의 DNA 단편의 크기를 결정하는 단계로서, 크기는 2개 이상의 크기 분포를 포함하는 것인 단계;

2개 이상의 크기 분포 중 각각의 크기 분포에 대하여

크기 분포 내에서 서열 태그에 대한 계층 위치를 결정하는 단계;

서열 태그를 피검자의 구성적 계층과 비교하여 제1 유전자좌의 첫 번째 수를 결정하는 단계로서, 제1 유전자좌 각각에서, 하나 이상의 서열 태그는 구성적 계층에 대해 서열 변이를 갖는 것인 단계;

제1 유전자좌에서 서열 변이를 갖는 서열 태그의 계층(count)을 기초로 하여 크기 분포에 대한 매개변수를 결정하는 단계로서, 이로써, 2개 이상의 크기 분포 중 제1 크기 분포에 대한 서열 변이의 제1 분획 농도를 제공하는 제1 매개변수 및 2개 이상의 크기 분포 중 제2 크기 분포에 대한 서열 변이의 제2 분획 농도를 제공하는 제2 매개변수가 결정되는 것인 단계; 및

제1 매개변수를 제2 매개변수와 비교하여 피검자에서 암의 수준의 분류를 결정하는 단계를 포함하는 방법.

청구항 2

제1항에 있어서, 다수의 DNA 단편의 크기를 결정하는 단계는,

피검자의 생물학적 시료 중의 다수의 DNA 단편의 크기 분획화를 수행하는 단계를 포함하는 방법.

청구항 3

제1항에 있어서, 다수의 DNA 단편 각각에 대하여, 하나 이상의 서열 태그는 DNA 단편의 양 말단을 포함하고, 다수의 DNA 단편의 크기를 결정하는 단계는,

양 말단에서 서열을 이용하여 DNA 단편의 크기를 결정하는 단계를 포함하는 방법.

청구항 4

제3항에 있어서, 하나 이상의 서열 태그는 쌍을 이룬 말단 서열 태그를 포함하고, 다수의 DNA 단편의 크기를 결정하는 단계는,

쌍을 이룬 말단 서열 태그를 기준 계층에 정렬하는 단계를 포함하는 방법.

청구항 5

제1항에 있어서, 다수의 DNA 단편의 크기를 결정하는 단계는,

다수의 DNA 단편 각각에 대한 하나 이상의 서열 태그로부터 다수의 DNA 단편의 크기를 결정하는 단계를 포함하고, 다수의 DNA 단편 각각에 대한 하나 이상의 서열 태그는 DNA 단편의 완전한 길이에 대한 서열 태그를 포함하는 방법.

청구항 6

제1항에 있어서, 제1 매개변수는 제1 크기 분포에 상응하고, 제2 매개변수는 제1 크기 분포보다 더 큰 제2 크기 분포에 상응하는 방법.

청구항 7

제6항에 있어서, 제1 매개변수가 제2 매개변수보다 더 높은 경우, 피검자에서 암의 수준은 비 혈액암으로 분류되는 방법.

청구항 8

제1항에 있어서, 크기 분포에 대한 매개변수는, 크기 분포에 대한 제1 유전자좌의 첫 번째 수에서 서열 변이를 나타내는 서열 태그의 합을 포함하는 방법.

청구항 9

제8항에 있어서, 합은 가중된 합이고, 제1 유전자좌 중 하나는 크기 분포에 대한 제1 가중치를 갖고, 이는 제1 유전자좌 중 두 번째가 갖는 크기 분포에 대한 제2 가중치와는 상이한 방법.

청구항 10

제9항에 있어서, 크기 분포에 대한 제1 가중치는 크기 분포에 대한 제2 가중치보다 더 크고, 제1 유전자좌 중 하나는 암과 관련되며, 제1 유전자좌의 두 번째는 암과 관련되지 않는 방법.

청구항 11

제1항에 있어서, 서열 태그에 대한 계층 위치를 결정하는 단계는,

서열 태그의 적어도 일부를 기준 계층에 정렬하는 단계를 포함하며, 서열 태그의 정렬이 서열 태그와 구성적 계층 사이의 하나 이상의 미스매치(mismatch)를 허용하는 방법.

청구항 12

제11항에 있어서, 각각의 크기 분포에 대한 서열 태그를 구성적 계층과 비교하는 단계는,

구성적 계층을 기준 계층과 비교하여 기준 계층에 대한 변이를 갖는 제2 유전자좌의 두 번째 수를 결정하는 단계;

정렬 단계를 기초로 하여, 크기 분포에 대한 제3 유전자좌의 세 번째 수를 결정하는 단계로서, 제3 유전자좌 각각에서, 하나 이상의 서열 태그는 기준 계층에 대해 서열 변이를 갖는 것인 단계; 및

세 번째 수와 두 번째 수의 차이를 구하여 크기 분포에 대한 제1 유전자좌의 첫 번째 수를 획득하는 단계를 포함하는 방법.

청구항 13

제12항에 있어서, 세 번째 수와 두 번째 수의 차이를 구하는 단계는, 제1 유전자좌의 구체적인 유전자좌를 확인하는 방법.

청구항 14

제13항에 있어서, 각각의 크기 분포에 대한 매개변수를 결정하는 단계는,

제1 유전자좌의 첫 번째 수의 각각의 유전자좌에 대하여, 유전자좌에 대해 정렬되고 유전자좌에서 서열 변이를 갖는 서열 태그를 계층하는 단계; 및

각각의 계층을 기초로 하여 크기 분포에 대한 매개변수를 결정하는 단계를 포함하는 방법.

청구항 15

제1항에 있어서, 구성적 계층은, 50% 초과와 구성적 DNA를 함유하는 피검자로부터의 구성적 시료로부터 기원하는 방법.

청구항 16

제1항에 있어서, 각각의 크기 분포에 대한 서열 태그에 대한 계층 위치를 결정하는 단계는,

서열 태그의 적어도 일부를 구성적 게놈에 정렬시키는 단계를 포함하며, 서열 태그의 정렬이 서열 태그와 구성적 게놈 사이의 하나 이상의 미스매치를 허용하는 방법.

청구항 17

제16항에 있어서, 각각의 크기 분포에 대한 서열 태그를 구성적 게놈과 비교하는 단계는,
정렬 단계를 기초로 하여, 피검자의 구성적 게놈에 대해 게놈 위치에서 서열 변이를 갖는 서열 태그를 확인하는 단계를 포함하며,
매개변수를 결정하는 단계는,
서열 변이를 나타내는 각각의 게놈 위치에 대하여,
게놈 위치에 대해 정렬되고 게놈 위치에서 서열 변이를 갖는 서열 태그의 각각의 첫 번째 수를 계수하는 단계;
및
각각의 첫 번째 수를 기초로 하여 크기 분포에 대한 매개변수를 결정하는 단계를 포함하는 방법.

청구항 18

제17항에 있어서, 각각의 첫 번째 수를 기초로 하여 크기 분포에 대한 매개변수를 결정하는 단계는,
각각의 첫 번째 수를 합하여 제1 합을 수득하는 단계; 및
제1 합을 사용하여 크기 분포에 대한 매개변수를 결정하는 단계를 포함하는 방법.

청구항 19

제18항에 있어서, 제1 합을 사용하여 크기 분포에 대한 매개변수를 결정하는 단계는,
제1 합으로부터 서열 변이를 나타내는 게놈 위치의 수를 감하는 단계를 포함하는 방법.

청구항 20

제18항에 있어서, 제1 합을 사용하여 크기 분포에 대한 매개변수를 결정하는 단계는,
정렬된 서열 태그의 양을 기초로 하여 제1 합을 정규화하는 단계를 포함하는 방법.

청구항 21

제1항에 있어서,
90% 초과와 구성적 DNA를 함유하는 피검자의 구성적 시료 중의 DNA 단편의 무작위 서열분석을 수행하여, 구성적 시료 중의 다수의 DNA 단편 각각에 대한 하나 이상의 제2 서열 태그를 수득하는 단계;
제2 서열 태그의 적어도 일부를 기준 게놈에 정렬하는 단계로서, 제2 서열 태그의 정렬은 M 이하의 게놈 위치에서 제2 서열 태그와 구성적 게놈 사이의 미스매치를 허용하며, 여기서 M은 1 이상의 정수인 것인 단계; 및
제2 서열 태그 및 정렬 단계를 기초로 하여 구성적 게놈을 제작하는 단계를 추가로 포함하는 방법.

청구항 22

제21항에 있어서, 구성적 시료는 생물학적 시료이고, 구성적 게놈을 제작하는 단계는,
2개의 대립형질을 가진 이형접합성 유전자좌 또는 동형접합성 유전자좌의 결정을 포함하는 공통 서열을 결정하는 단계; 및
구성적 게놈으로 공통 서열을 사용하는 단계를 포함하는 방법.

청구항 23

제1항에 있어서,
피검자의 생물학적 시료 중의 DNA 단편의 각각의 크기 분포의 무작위 서열분석을 수행하여, 생물학적 시료 중의

DNA 단편의 각각의 크기 분포 중 다수의 DNA 단편 각각에 대한 하나 이상의 서열 태그를 생성하는 단계를 추가로 포함하는 방법.

청구항 24

제1항에 있어서, 하나 이상의 서열 태그는, 생물학적 시료 중의 DNA 단편의 각각의 크기 분포 중 DNA 단편을 무작위 서열분석하는 것으로부터 생성되는 방법.

청구항 25

제1항에 있어서, 생물학적 시료는 뇨, 흉막액, 복수액, 복강액, 타액, 뇌척수액, 또는 대변 시료인 방법.

청구항 26

제1항에 있어서, 제1 매개변수는 제1 크기 분포의 종양 유래 DNA의 제1 분획 농도이고, 제2 매개변수는 제2 크기 분포의 종양 유래 DNA의 제2 분획 농도인 방법.

청구항 27

실행시 제1항 내지 제26항 어느 한 항의 방법을 수행하도록 컴퓨터 시스템을 제어하는 다수의 명령을 저장하는 비전송적 컴퓨터 판독가능한 매체.

청구항 28

제27항의 비전송적 컴퓨터 판독가능한 매체를 포함하는 컴퓨터 시스템.

발명의 설명

기술 분야

[0001] 관련 출원의 상호 참조

[0002] 본 출원은 모든 목적을 위해 이의 전문이 본원에 참조로 혼입된, 2012년 6월 21일자로 출원된 "암 검출용 혈장 DNA의 돌연변이 분석"이라는 명칭의 미국 가특허원 제61/662,878호; 2012년 8월 13일자로 출원된 "암 검출용 혈장 DNA의 돌연변이 분석"이라는 명칭의 미국 가특허원 제61/682,725호; 2012년 8월 31일자로 출원된 "암 검출용 혈장 DNA의 돌연변이 분석"이라는 명칭의 미국 가특허원 제61/695,795호; 및 2012년 10월 8일자로 출원된 "암 검출용 혈장 DNA의 돌연변이 분석"이라는 명칭의 미국 가특허원 제61/711,172호의 비-가특허원이며 이들의 이익을 청구한다.

배경 기술

[0003] 종양-기원한 DNA는 암 환자의 무세포 혈장/혈청에 존재하는 것으로 밝혀져 왔다(참조: Chen XQ 등 Nat Med 1996; 2: 1033-1035). 가장 흔한 방법은 암과 관련된 것으로 공지된 돌연변이의 직접 분석을 기준으로 한다(참조: Diehl F 등 Proc Natl Acad Sci 2005; 102: 16368-16373; Forsheew T 등 Sci Transl Med 2012; 4: 136ra68). 다른 방법은 혈장 DNA의 무작위적인 서열분석으로 검출된 암-관련된 카피 수 변이를 시험하여 왔다(Lo 등의 미국 특허 공개공보 제2013/0040824호).

[0004] 시간이 흘러, 하나 이상이 암 세포가 성장 장점을 획득하여 딸 세포의 다수의 클론을 생산할 수 있음이 알려져 있다. 궁극적으로, 종양 성장 및/또는 이의 전이 초점은 클론성 암 세포의 그룹의 집합체를 함유할 수 있다. 이러한 현상은 전형적으로 종양 이종성(heterogeneity)으로 언급된다(참조: Gerlinger M 등 N Engl J Med 2012; 366: 883-892; Yap TA 등 Sci Transl Med 2012; 4: 127ps1 0).

[0005] 암은 고도로 이종성이며, 즉 동일한 조직 유형의 암의 돌연변이 프로파일은 광범위하게 변할 수 있음이 알려져 있다. 따라서, 특이적인 돌연변이의 직접적인 분석은 전형적으로 이들 특이적인 돌연변이와 관련된 것으로 공지된 특수 암 유형 내에서 경우들의 서브세트 만을 검출할 수 있다. 또한, 종양-기원한 DNA는 일반적으로 사람 혈장 내 DNA의 약간의 중이며; 혈장 속의 DNA의 절대적인 농도는 낮다. 따라서, 혈장 또는 혈청 속의 암-관련된 돌연변이 중의 하나 또는 소그룹의 직접적인 검출은 표적화된 돌연변이를 지닌 것으로 공지된 암을 지닌 환자들 중에서조차 낮은 분석적 민감성을 달성할 수 있다. 더욱이, 단일 종양 내에서조차 돌연변이의 측면에서 유의적인 종양 내 이종성이 존재함이 밝혀져 왔다. 이러한 돌연변이는 종양 세포의 소집단에서만 찾을 수

있다. 원발성 종양과 전이성 병변 사이의 돌연변이 프로파일에 있어서의 차이는 훨씬 더 크다. 종양 내 및 원시-전이 이중성의 한가지 예는 결장직장 암으로 고생하는 환자에서 KRAS, BRAF 및 PIK3CA 유전자를 포함한다(참조: Baldus 등 Clin Cancer Research 2010, 16:790-9).

[0006] 환자가 원시 종양(KRAS 돌연변이를 수반하나 PIK3CA 돌연변이는 아님) 및 감춰진 전이성 병변(PIK3CA 돌연변이를 포함하나 KRAS 돌연변이는 아님)을 갖는 시나리오에 있어서, 원시 종양 내 KRAS 돌연변이의 검출에 초점을 맞추는 경우, 감춰진 전이성 병변은 검출될 수 없다. 그러나, 분석시 돌연변이 둘 다를 포함하는 경우, 원시 종양 및 감춰진 전이성 병변 둘 다를 검출될 수 있다. 따라서, 돌연변이 둘 다를 포함하는 시험은 잔류성 종양 조직의 검출시 더 높은 민감성을 가질 수 있다. 이러한 단순한 예는, 암에 대해 스크리닝하는 경우, 및 발생할 수 있는 돌연변이 유형이 거의 없거나 이의 단서가 없으므로 보다 복잡해질 수 있다.

[0007] 따라서, 암의 광범위한 스크리닝, 검출, 또는 평가를 수행하기 위한 새로운 기술을 제공하는 것이 바람직할 수 있다.

발명의 내용

[0008] 요약

[0009] 구현예들은 암을 스크리닝하거나 모니터링하는 피검자의 구성적 DNA 내의 돌연변이와 비교하여, 동일한 피검자의 생물학적 시료(예를 들면, 혈장 또는 혈청)내 체세포 돌연변이의 빈도를 관찰할 수 있다. 무작위적인 서열 분석을 사용하여 이들 빈도를 측정할 수 있다. 매개변수는 이들 빈도로부터 기원할 수 있으며 암의 수준의 분류를 결정하는데 사용할 수 있다. 거짓 양성은 규정된 수의 변이체 서열 판독[태그(tag)]을 갖도록 하는 특성의 변이체 유전자좌(locus)를 요구함으로써 보다 정밀한 매개변수를 제공함에 의해 필터링하여 제거될 수 있다. 상이한 변이체 유전자좌에 대한 상대적인 빈도를 분석하여 환자내 종양의 이중성의 수준을 측정할 수 있다.

[0010] 하나의 구현예에서, 매개변수는 암이 없거나, 암 위험이 낮은 피검자의 그룹으로부터 기원한 동일한 매개변수와 비교할 수 있다. 시험 피검자로부터 수득된 매개변수 및 암이 없거나, 암 위험이 낮은 피검자의 그룹으로부터의 매개변수에 있어서의 유의적인 차이는, 시험 피검자가 암 또는 전암성(premalignant) 상태이거나 장래에 암으로 발달할 수 있는 위험이 증가되어 있음을 나타낼 수 있다. 따라서, 하나의 구현예에서, 혈장 DNA 분석은 종양의 사전 계층 정보없이 수행할 수 있다. 따라서, 이러한 구현예는 암을 스크리닝하는데 특히 유용하다.

[0011] 다른 구현예에서, 구현예는 또한 치료 후 암 환자를 모니터링하여 잔류 종양이 존재하는지 또는 종양이 재발하였는지를 알아보는데 사용할 수 있다. 예를 들면, 잔류 암을 지니거나, 종양이 재발된 환자는 잔류 종양이 존재하지 않거나 종양 재발이 관찰되지 않는 환자보다 체세포 돌연변이의 빈도가 더 높을 수 있다. 모니터링은 암 환자로부터 체액 또는 무세포 핵산을 지닌 다른 시료, 예를 들면, 혈장 또는 혈청 속에서 종양-관련된 유전적 이탈의 일시적 변화를 추정하기 위해 치료 후 다수의 시점에서 암 환자로부터의 시료를 수득함을 포함할 수 있다.

[0012] 하나의 구현예에 따라서, 방법은 피검자에서 암 또는 전암성 변화를 검출한다. 피검자의 구성적 게놈이 수득된다. 하나 이상의 서열 태그가 피검자의 생물학적 시료 속의 다수의 DNA 단편 각각에 대해 수용되며, 여기서 생물학적 시료는 세포-유리된 DNA를 포함한다. 게놈 위치는 서열 태그에 대해 측정된다. 서열 태그는 구성적 게놈과 비교하여 제1 유전자 위치의 첫 번째 수를 결정한다. 각각의 제1 유전자 위치에서, 구성적 게놈과 비교하여 서열 변이를 갖는 서열 태그의 수는 컷오프(cutoff) 값을 초과하며, 여기서 컷오프 값은 1보다 더 크다. 매개변수는 제1 유전자좌에서 서열 변이를 갖는 서열 태그의 수를 기초로 하여 결정한다. 매개변수는 역치 값(threshold value)과 비교하여 상기 피검자에서 암의 수준의 분류를 측정한다.

[0013] 다른 구현예에 따라서, 방법은 피검자의 하나 이상의 종양의 이중성을 분석한다. 피검자의 구성적 게놈을 수득한다. 하나 이상의 서열 태그는 피검자의 생물학적 시료 속의 다수의 DNA 단편 각각에 대해 수용되며, 여기서 생물학적 시료는 무세포 DNA를 포함한다. 게놈 위치는 서열 태그에 대해 측정한다. 서열 태그를 구성적 게놈과 비교하여 제1 유전자 위치의 첫 번째 수를 결정한다. 제1 유전자 위치 각각에서, 구성적 게놈과 비교하여 서열 변이를 갖는 서열 태그의 수는 컷오프 값 초과이며, 여기서 당해 컷오프 값은 1보다 크다. 하나 이상의 종양의 이중성의 척도는 제1 게놈 위치의 세트의 각각의 첫 번째 수를 기초로 계산한다.

[0014] 다른 구현예에 따라서, 방법은 세포-유리된 DNA를 포함하는 생물학적 시료 중의 종양 DNA의 분획 농도를 측정한다. 하나 이상의 서열 태그는 생물학적 시료 중의 다수의 DNA 단편 각각에 대해 수용된다. 게놈 위치는 서열 태그에 대해 측정된다. 다수의 게놈 영역 각각에 대해, 게놈 영역 내 DNA 단편의 각각의 양을 게놈 영역 내 게

놈 위치를 갖는 서열 태그로부터 측정한다. 각각의 양을 정규화하여 각각의 밀도를 수득한다. 각각의 밀도는 기준 밀도와 비교하여 게놈 영역이 1개-카피 손실 또는 1개-카피 획득을 나타내는지를 확인한다. 제1 밀도는 1개 카피 획득을 나타내는 것으로서 확인된 각각의 밀도로부터 또는 1개 카피 손실을 나타내는 것으로 확인된 각각의 밀도로부터 계산한다. 분획 농도는 제1 밀도를 다른 밀도와 비교하여 차이를 수득함으로써 계산하며, 여기서 차이는 기준 밀도를 사용하여 정규화한다.

[0015] 다른 구현예는 본원에 기술된 방법과 관련된 시스템 및 컴퓨터 판독가능한 매체에 관한 것이다.

[0016] 본 발명의 특성 및 장점의 보다 나은 이해는 다음의 상세한 설명 및 첨부된 도면을 참조하여 수득될 수 있다.

도면의 간단한 설명

[0017] 도 1은 본 발명의 구현예에 따라서 피검자에서 암 또는 전암성 변화를 검출하기 위한 방법 (100)의 흐름도이다.

도 2는 본 발명의 구현예에 따라서 시료 게놈(SG)을 구성적 게놈(CG)과 직접 비교하는 방법의 흐름도를 나타낸다.

도 3은 본 발명의 구현예에 따라서 기준 게놈(RG)을 사용하여 시료 게놈(SG)을 구성적 게놈(CG)과 비교하는 방법 (300)의 흐름도를 나타낸다.

도 4는, 시료 속에서 종양-기원한 DNA의 분획 농도가 10%인 것으로 추정되는 경우 본 발명의 구현예에 따라 시료 속에 존재하는 것으로 돌연변이를 분류하기 위한 기준으로 상이한 발생수를 사용하여 정확하게 확인된 암-관련 단일 뉴클레오타이드 돌연변이의 수를 나타내는 표 (400)이다.

도 5는, 시료 속의 종양-기원한 DNA의 분획 농도가 5%인 것으로 추정되는 경우 확인된 돌연변이의 예측된 수 및 위-양성 유전자좌의 예측된 수를 나타내는 표이다.

도 6a는 종양-기원한 DNA의 10% 및 20% 혈장 분획 농도를 사용한 혈장 속의 암-관련 돌연변이의 검출률을 나타내고 잠재적인 암-관련 돌연변이를 묘사하기 위한 기준으로서 4개 및 6개의 발생(r)을 사용하는 그래프 (600)이다. 도 6b는 4, 5, 6 및 7의 발생(r) 기준 대 서열분석 깊이를 사용하여 뉴클레오타이드 변화를 갖는 것으로 거짓 분류된 뉴클레오타이드 위치의 예측된 수를 나타내는 그래프 (650)이다.

도 7a는, 시료 속에서 종양-기원한 DNA의 분획 농도가 5%인 것으로 추정되는 경우 차이 서열분석 깊이를 사용하여 암-관련 돌연변이 부위 및 거짓 양성 부위의 수를 나타내는 그래프 (700)이다. 도 7b는 전체 게놈(WG) 및 모든 엑손(exon)의 분석을 포함하는 거짓 양성 부위의 예측된 수를 나타내는 그래프 (750)이다.

도 8은 본 발명의 구현예에 따라 혈장 속에서 종양-기원한 DNA의 분획 농도를 포함하는, 치료 전 및 후에 4개의 HCC 환자에 대한 결과를 나타내는 표 (800)이다.

도 9는 본 발명의 구현예에 따라서 16명의 건강한 대조군 피검자에서 HCC-관련된 SNV의 검출을 나타내는 표 (900)이다.

도 10a는 본 발명의 구현예에 따라서 HCC 환자의 종양 시료의 서열 판독 밀도의 분포 플롯을 나타낸다. 도 10b는 본 발명의 구현예에 따라서 HCC 환자의 혈장 속에서 모든 빈(bin)에 대한 z-점수의 분포 플롯 (1050)을 나타낸다.

도 11은 본 발명의 구현예에 따라서 HCC 환자의 혈장에 대한 z-점수의 분포 플롯 (1100)을 나타낸다.

도 12는 본 발명의 구현예에 따라서 세포-유리된 DNA를 포함하는 생물학적 시료 중의 종양 DNA의 분획 농도를 결정하는 방법 (1200)의 흐름도이다.

도 13a는 본 발명의 구현예에 따라서 진단 시기에 유방 암 및 난소 암을 지닌 환자의 혈장 속에서 돌연변이의 분석의 표 (1300)을 나타낸다.

도 13b는 본 발명의 구현예에 따라서 종양 절개 후 쌍방 난소 암 및 유방 암을 지닌 환자의 혈장 속에서 돌연변이의 분석의 표 (1350)을 나타낸다.

도 14a는 HCC1에 대한 혈장 DNA 속에서 단일 뉴클레오타이드 변화의 검출을 나타내는 표 (1400)이다. 도 14b는 HCC2에 대한 혈장 DNA 속에서 단일 뉴클레오타이드 변화의 검출을 나타내는 표 (1450)이다.

도 15a는 HCC3에 대한 혈장 DNA 속에서 단일 뉴클레오타이드 변화의 검출을 나타내는 표 (1500)이다. 도 15b는

HCC4에 대한 혈장 DNA 속에서 단일 뉴클레오타이드 변화의 검출을 나타내는 표 (1550)이다.

도 16은 난소(및 유방) 암 환자에 대한 혈장 DNA 속에서 단일 뉴클레오타이드 변화의 검출을 나타내는 표 (1600)이다.

도 17은 발생 및 서열분석 깊이의 상이한 요건의 예측된 민감성을 나타내는 표 (1700)이다.

도 18은 상이한 컷오프 및 상이한 서열분석 깊이에 대한 거짓 양성 유전자좌의 예측된 수를 나타내는 표 (1800)이다.

도 19는 상이한 종양 부위에서 검출된 돌연변이의 수를 나열하는 수형도를 나타낸다.

도 20은 치료 전 및 치료 후 혈장 시료 속에서 종양-기원한 돌연변이를 수반하는 단편의 수를 나타내는 표 (2000)이다.

도 21은 모든 4개의 종양 부위에서 검출된 단일의 종양 부위 및 돌연변이에서 검출된 돌연변이에 대한 혈장 속의 발생의 분포를 나타내는 그래프 (2100)이다.

도 22는 이중 종양으로부터 온 돌연변이에 대한 혈장 내 발생의 예측된 분포를 나타내는 그래프 (2200)이다.

도 23은 모집된 16명의 건강한 대조군 피검자에 대한 구현예의 특이성을 입증한다.

도 24는 본 발명의 구현예에 따른 피검자의 하나 이상의 종양의 이중성을 분석하기 위한 방법 (2400)의 흐름도이다.

도 25는 본 발명의 구현예에 따른 시스템 및 방법으로 사용가능한 컴퓨터 시스템 (2500) 예의 블록 선도이다.

발명을 실시하기 위한 구체적인 내용

[0018] 정의

[0019] 본원에 사용된 것으로서, 용어 "유전자좌" 또는 이의 복수형 "유전자좌들"은 게놈에 걸친 변이를 가질 수 있는 뉴클레오타이드(또는 염기쌍)의 특정 길이의 위치 또는 주소이다. "빈(bin)"은 게놈 내 예정된 길이의 영역이다. 빈의 수는 동일한 첫 번째 길이(분해도)이지만, 상이한 수는 동일한 제2 길이를 가질 수 있다. 하나의 구현예에서, 빈은 서로 중첩되지 않는다.

[0020] 본원에 사용된 것으로서, 용어 "무작위적 서열분석"은, 서열분석된 핵산 단편이 서열분석 과정 전에 구체적으로 확인되지 않거나 예정되지 않은 서열분석을 말한다. 특이적인 유전자 유전자좌를 표적화하기 위한 서열-특이적인 프라이머는 요구되지 않는다. 용어 "공통의 서열분석(universal sequencing)"은, 서열분석이 특정 단편에서 출발할 수 있는 서열분석을 말한다. 하나의 구현예에서, 어댑터(adapter)를 단편의 말단에 가하고, 서열분석을 위한 프라이머를 어댑터에 부착한다. 따라서, 어떠한 단편도 동일한 프라이머로 서열분석될 수 있으므로, 서열분석은 무작위일 수 있다.

[0021] 본원에 사용된 것으로서 용어 "서열 태그"는 핵산 분자의 특정 부분 또는 모두로부터 서열분석된 뉴클레오타이드의 스트링(string)을 말한다. 예를 들어, 서열분석된 태그는 핵산 단편으로부터 서열분석된 뉴클레오타이드(예를 들면, ~30)의 짧은 스트링, 핵산 단편의 양쪽 말단에서 뉴클레오타이드의 짧은 스트링, 또는 생물학적 시료 속에 존재하는 전체 핵산 단편의 서열 분석일 수 있다. 핵산 단편은 보다 큰 핵산 분자의 특정 부분이다. 단편(예를 들어, 유전자)는 보다 큰 핵산 분자의 다른 부분에 대해 별도로(즉, 연결되지 않고) 존재할 수 있다.

[0022] 용어 "구성적 게놈"(또한 CG로 언급됨)은 게놈 내 유전자 자리에서 컨센서스 서열(consensus sequence)로 구성된다. CG는 피검자의 전체 게놈(예를 들면, 사람 게놈), 또는 게놈의 바로 일부를 포함할 수 있다. 구성적 게놈(CG)은 세포의 DNA 및 또한 세포-유리된 DNA(예를 들면, 혈장 속에서 발견될 수 있는 것으로서)로부터 수득될 수 있다. 이상적으로, 컨센서스 뉴클레오타이드는, 유전자좌가 1개의 대립형질에 대해 동형접합체이거나 2가지 대립형질에 대해 이형접합체이다. 이형접합체 유전자좌는 전형적으로 유전 다형체의 구성원인 2가지 대립형질을 함유한다. 예로서, 유전자좌가 이형접합체인지를 결정하는 기준은 유전자좌에 대해 정렬된 판독물의 적어도 예정된 퍼센트(예를 들면, 30% 또는 40%)에서 각각 나타나는 2가지 대립형질의 역치일 수 있다. 하나의 뉴클레오타이드가 충분한 퍼센트(예를 들면, 70% 초과)로 나타나는 경우, 유전자좌는 CG에서 동형접합체인 것으로 추정될 수 있다. 하나의 건강한 세포의 게놈이 세포 분열 동안에 자발적으로 발생하는 무작위적 돌연변이로 인해 다른 건강한 세포의 게놈과는 상이할 수 있다고 해도, 이러한 컨센서스가 사용되는 경우 CG는 변하지 않아야 한다. 일부 세포, 예를 들어, 항체 및 T 세포 수용체 유전자를 포함하는 것과 같이 B 및 T 림프구는 게놈 재배열

을 지닌 게놈을 지닐 수 있다. 이러한 거대 규모 차이는 여전히 혈액 속에서 전체의 핵화된 세포 집단의 비교적 소 집단일 수 있으므로, 이러한 재배열은 혈액 세포의 충분한 시료채취(예를 들면, 서열분석 깊이)로 구성적 게놈의 측정에 영향을 미치지 않을 수 있다. 볼 세포, 피부 세포, 모낭을 포함하는 다른 세포 유형, 또는 각종 정상 체 조직의 생검은 또한 CG의 공급원으로서 제공될 수 있다.

[0023] 용어 "구성적 DNA"는, 피검자가 출생한 유전자구성의 반영인 DNA의 어떠한 공급원을 말한다. 피검자의 경우, 구성적 DNA가 수득될 수 있는 "구성적 시료"의 예는 건강한 혈액 세포 DNA, 볼 세포 DNA 및 모근 DNA를 포함한다. 이들 건강한 세포로부터의 DNA는 피검자의 CG를 정의한다. 이들 세포는 예를 들면, 개인이 암을 가지지 않은 것으로 알려진 경우 또는 시료가 암성 또는 악성 세포(예를 들면, 간암이 추정되는 경우 모근 DNA)를 함유하지 않을 것 같은 조직으로부터 수득될 수 있는 경우, 다양한 방법으로 건강한 것으로 확인될 수 있다. 또 다른 예로서, 혈장 시료는, 환자가 암이 없는 경우 수득하여, 측정된 구성적 DNA를 후속적인 혈장 시료(예를 들면, 1년 이상 후)로부터의 결과에 대해 비교할 수 있다. 다른 구현예에서, <50%의 종양 DNA를 함유하는 단일의 생물학적 시료는 구성적 게놈 및 종양-관련된 유전적 변형을 추론하기 위해 사용될 수 있다. 이러한 시료에서, 종양-관련된 단일의 뉴클레오타이드 돌연변이의 농도는 CG에서 이형접합성 SNP의 각각의 대립형질의 것보다 더 낮을 수 있다. 이러한 시료는 하기 기술한, 시료 게놈을 결정하는데 사용된 생물학적 시료와 동일할 수 있다.

[0024] 본원에 사용된 것으로서 용어 "생물학적 시료"는 피검자(예를 들면, 사람, 암을 지닌 사람, 암을 지닌 것으로 추측되는 사람, 또는 다른 유기체)로부터 취한 특정 시료를 말하며 목적하는 하나 이상의 세포-유리된 핵산 분자(들)를 함유한다. 생물학적 시료는 세포-유리된 DNA를 포함할 수 있으며, 이들 중 일부는 건강한 세포로부터 기원할 수 있고 일부는 종양 세포로부터 기원할 수 있다. 예를 들어, 종양 DNA는 혈액 또는 다른 유액, 예를 들면, 뇨, 흉막액, 복수액, 복강액, 타액, 눈물 또는 뇌척수액에서 발견될 수 있다. 비-유액의 예는 대변 시료이며, 이는 설사액과 혼합될 수 있다. 이러한 시료 중 일부의 경우, 생물학적 시료는 비-침입적으로 수득될 수 있다. 일부 구현예에서, 생물학적 시료는 구성적 시료로서 사용될 수 있다.

[0025] 용어 "시료 게놈"(또는 SG로 언급됨)은 게놈(예를 들면, 사람 게놈)의 위치에 정렬되어진 서열 판독물의 수집이다. 시료 게놈(SG)은 공통 서열이 아니나, 충분한 수의 판독물(예를 들면, 적어도 2 또는 3, 또는 그 이상의 컷오프 값)만으로 나타낼 수 있는 뉴클레오타이드를 포함한다. 대립형질이 충분한 횟수로 나타나고 CG의 일부가 아닌 경우(즉, 공통 서열의 일부가 아닌 경우), 이러한 대립형질은 "단일 뉴클레오타이드 돌연변이"를 나타낼 수 있다(또한 SNM으로 언급됨). 다른 유형의 돌연변이, 예를 들면, 2개 이상의 뉴클레오타이드를 포함하는 돌연변이(예를 들면, 미소부수체로서 또는 단일의 탠덤 반복 다형체(tandem repeat polymorphism)로서 탠덤 반복 단위의 수에 영향을 미침), 염색체 전좌(이는 염색체내 또는 염색체간일 수 있다) 및 서열 역전을 또한 본 발명을 사용하여 검출할 수 있다.

[0026] 용어 "기준 게놈"(또한 RG로 언급됨)은, 이에 대해 서열이 생물학적 시료로부터 판독되고 구성적 시료가 정렬되어 비교될 수 있는 반수체(haploid) 또는 이배체 게놈을 말한다. 반수체 게놈의 경우, 각각의 유전자좌에 단지 하나의 뉴클레오타이드가 존재한다. 이배체 게놈의 경우, 이형접합체 유전자좌가 확인될 수 있으며, 이러한 유전자좌는 2가지 대립형질을 가지고, 여기서 대립형질은 유전자좌에 정렬하기 위해 조화를 허용할 수 있다.

[0027] 용어 "암의 수준"은, 암이 존재하는지의 여부, 암의 단계, 종양의 크기, 및/또는 암의 중증도의 다른 척도를 말할 수 있다. 암의 수준은 수 또는 다른 특성일 수 있다. 당해 수준을 0일 수 있다. 암의 수준은 또한 돌연변이 또는 돌연변이의 수와 관련된 악성 또는 전암성 조건(상태)를 포함한다. 암의 수준은 각종 방법으로 사용될 수 있다. 예를 들어, 암을 이미 지닌 것으로 알려져 있지 않은 사람에서 암이 존재하는 경우, 스크리닝을 점검할 수 있다. 평가는 암으로 진단된 사람을 시험할 수 있다. 검출은 '스크리닝'을 의미하거나 암의 제안된 특징(예를 들면, 증상 또는 다른 양성 시험)을 지닌 사람이 암을 지니고 있는지를 점검함을 의미할 수 있다.

[0028] 상세한 설명

[0029] 구현예는 종양으로부터 직접 취하지 않고 세포-유리된 핵산을 포함하는 생물학적 시료(예를 들면, 혈액 혈장/혈청 시료)의 분석에 의한 암의 검출을 위해 제공된다. 세포-유리된 핵산은 신체 전체에서 조직의 다양한 유형에 대해 기원할 수 있다. 이러한 방식으로, 각종 암의 검출을 위한 광범위한 분석을 수행할 수 있다.

[0030] 유전적 일탈(단일의 뉴클레오타이드 돌연변이, 검출, 증폭, 및 재배열 포함)은 암의 발달 동안에 종양 세포속에 축적된다. 구현예에서, 거대한 평행 서열분석을 사용하여 체액(예를 들면, 혈장, 혈청, 타액, 복수액, 활막액 및 뇌척수액) 속에서 또한 단일 뉴클레오타이드 변이(SNV)로 불리는, 단일 뉴클레오타이드 돌연변이(SNM)를 검

출하고 정량함으로써 암을 검출하고 모니터링할 수 있다. SNM(또는 다른 유형의 돌연변이)의 수의 정량화는 스크리닝 시험의 일부로서 암의 초기 단계를 확인하기 위한 메커니즘을 제공할 수 있다. 각종 시행에서, 서열분석 오차를 구별하고 건강한 세포에서 발생하는(예를 들면, 특수 유전자 자리에서 확인될 SNM의 수, 예를 들면, 적어도 3, 4 또는 5개의 SNM을 필요로 함으로써) 자발적인 돌연변이를 구별하기 위해 주의를 기울인다.

[0031] 일부 양태는 또한 종양 이중성의 분석을 위한 비침습성 방법을 제공하며, 이는 동일한 종양(즉, 종양 내 이중성)내 세포 또는 신체내 상이한 종양(동일한 부위 또는 상이한 부위로부터)으로부터의 세포를 포함할 수 있다. 예를 들어, 각각의 돌연변이를 함유하는 상대적인 종양 세포 덩어리의 평가를 포함하는, 이러한 종양 이중성의 클론 구조를 비침습적으로 분석할 수 있다. 비교적 고 농도로 존재하는 돌연변이는 신체내 다수의 악성 세포, 예를 들면, 신체내 여전히 다른 악성 세포에 대해 종양생성 공정 동안에 초기 발생한 세포 내에 존재한다(참조: Welch JS 등 Cell 2012; 150: 264-278). 이러한 돌연변이는, 이들의 상대적으로 보다 높은 풍부성으로 인하여, 비교적 보다 낮은 풍부성을 지닌 것보다 암 DNA를 검출하기 위한 보다 높은 진단적 민감성을 나타내는 것으로 예측된다. 돌연변이의 상대적인 풍부성의 변화의 일련의 모니터링은, 질병이 진행하면서 자발적으로, 또는 치료에 대한 반응시, 종양의 클론 구조에 있어서의 변화를 비침습적으로 모니터링하도록 한다. 이러한 정보는 치료에 대한 종양 내성의 초기 검출시 또는 예후의 평가시 사용될 수 있다.

[0032] I. 도입

[0033] 돌연변이는 DNA 복제 및/또는 DNA 보수시 오차로 인하여 세포 분열동안 발생할 수 있다. 이러한 돌연변이들 중 한가지 유형은 단일 뉴클레오타이드의 변경을 포함하며, 이는 게놈의 상이한 부분으로부터 다수의 서열을 포함할 수 있다. 암은 일반적으로 성장 장점을 획득한 단일 암 세포의 클론적 확장에 기인한 것으로 여겨진다. 이러한 클론 확장은 조상 암 세포로부터 기원하는 모든 암 세포 내 돌연변이(예를 들면, 단일의 뉴클레오타이드 돌연변이)의 축적을 초래할 수 있다. 이러한 후대 종양 세포는 돌연변이 세트(예를 들면, 단일의 뉴클레오타이드 돌연변이)를 공유할 수 있다. 본원에 기술된 바와 같이, 암-관련된 단일의 뉴클레오타이드 돌연변이는 암 환자의 혈장/혈청 속에서 검출될 수 있다.

[0034] 일부 양태는 생물학적 시료(예를 들면, 혈장 또는 혈청) 속에서 모든 돌연변이를 효과적으로 선별할 수 있다. 돌연변이의 수가 고정되어 있지 않으므로(종양 세포의 상이한 소집단으로부터 수백, 수천 또는 수백만의 암-관련 돌연변이가 검출될 수 있다), 구현에는 특이적인 돌연변이를 검출하는 기술보다 더 우수한 민감성을 제공할 수 있다. 돌연변이의 수는 암을 검출하는데 사용될 수 있다.

[0035] 많은 또는 모든 돌연변이의 이러한 스크리닝을 제공하기 위하여, 종양-기원한 DNA를 함유할 수 있는 생물학적 시료(예를 들면, 혈장 및 혈청을 포함하는 체액) 속에서 유전적 변화에 대한 조사(예를 들면, 무작위 조사)를 수행할 수 있다. 혈장과 같은 시료의 사용은 종양 또는 암의 침습적 생검을 수행할 필요성을 제거한다. 또한, 스크리닝은 게놈의 모든 또는 거대한 영역을 포함할 수 있으므로, 이러한 스크리닝은 어떠한 거대하고 알려진 돌연변이에만 한정되는 것이 아니라, 어떠한 돌연변이의 존재도 사용할 수 있다. 더욱이, 돌연변이의 수는 게놈의 모든 또는 거대한 영역에 걸쳐 합해지므로, 보다 높은 민감성이 획득될 수 있다.

[0036] 그러나, 돌연변이로 계수되지 않아야 하는, 사람 게놈 내 단일의 뉴클레오타이드 다형성(SNP)을 포함하는, 다형성 부위가 존재한다. 구현예들은, 검출되어진 유전적 변이가 암-관련된 돌연변이인지 또는 게놈 내 다형성인지를 확인할 수 있다. 예를 들어, 게놈 내 암-관련된 돌연변이와 다형성 사이를 결정하는 부분으로서, 구현예는 다형성을 포함할 수 있는, 구성적 게놈을 측정할 수 있다. 구성적 게놈(CG)의 다형성은 서열분석 데이터 내 충분히 높은 퍼센트(예를 들면, 30 내지 40%)로 나타난 다형성으로 제한될 수 있다.

[0037] 이후에, 생물학적 시료로부터 수득된 서열은 단일의 뉴클레오타이드 돌연변이(SNM)이거나, 확인된 다른 유형의 돌연변이인 구성적 게놈 및 변이에 정렬될 수 있다. 이들 SNM은 공지된 다형성내 포함되지 않은 변이일 수 있으므로, 암-관련된, 및 구성적 게놈의 부분이 아닌 것으로 표지될 수 있다. 건강한 사람은 예를 들면, 세포 분열 동안에 생성된, 건강한 세포 중에서 무작위 돌연변이로 인하여 특정 수의 SNM을 가질 수 있지만, 암 환자는 보다 많은 SNM을 가질 수 있다.

[0038] 예를 들어, 암 환자의 경우, 체액 속에서 검출가능한 SNM의 수는 동일한 사람의 구성적 게놈 속에 존재하는 다형성보다 더 높을 수 있다. 종양-기원한 DNA와 대부분의 구성적 DNA를 함유하는 DNA 시료를 함유하는 체액 시료 속에서 검출된 변이의 양 사이에서 비교를 수행할 수 있다. 하나의 구현예에서, 용어 '대부분'은 90% 초과를 의미할 수 있다. 다른 바람직한 구현예에서, 용어 '대부분'은 95%, 97%, 98%, 또는 99%를 의미할 수 있다. 체액 속의 변이의 양이 대부분 구성적 DNA를 지닌 시료의 것을 초과하는 경우, 체액이 종양-기원한 DNA를 함유

할 수 있는 확률이 증가한다.

[0039] DNA 시료 속에서 변이를 무작위적으로 조사하는데 사용될 수 있는 한가지 방법은 무작위 또는 셋건 서열분석(shotgun sequencing)(예를 들면, 거대한 평행 서열분석 사용)이다. 연결에 의한 서열분석 플랫폼(sequencing-by-ligation platform)(예를 들면, Life Technologies SOLiD 플랫폼), 이온 토런트(Ion Torrent)/이온 프로톤(Ion Proton), 반도체 서열분석, Roche 454, 단일 분자 서열분석 플랫폼(예를 들면, Helicos, Pacific Biosciences 및 nanopore)을 포함하는 어떠한 거대한 평행 서열분석 플랫폼도 사용할 수 있다. 여전히, 서열분석 오차가 발생할 수 있으며 구성적 DNA내 변이로서 또는 종양 DNA로부터 기원한 돌연변이로 잘못 해석될 수 있음이 알려져 있다. 따라서, 본 발명자의 제안된 시도의 특이성을 개선시키기 위하여, 서열분석 오차 또는 분석적 오차의 다른 성분들의 확률을 예를 들면, 이것이 SNM으로서 계수될 유전자좌에서 검출된 대립형질의 적어도 규정된 수(예를 들면, 2 또는 3)를 필요로하면서 적절한 서열분석 깊이를 사용하여 계수할 수 있다.

[0040] 본원에 기술된 것으로서, 구현예는, 시료 속에 존재하는 무작위적으로 검출된 유전적 변이의 양이 분석적 오차(예를 들면 서열분석 오차)로 인하여 부적절하게 검출될 수 있는 구성적 DNA 및 변이에 대해 예측된 것을 초과하는 경우 생물학적 시료(예를 들면, 체액)속에서 종양-기원한 DNA의 존재에 대한 증거를 제공할 수 있다. 당해 정보는 암의 스크리닝, 진단, 예후 및 모니터링에 사용될 수 있다. 다음의 단락에서, 본 발명자들은 혈장/혈청 또는 다른 시료(예를 들면, 체액)속에서 단일의 뉴클레오타이드 돌연변이의 검출을 위해 사용될 수 있는 분석 단계를 기술한다. 체액은 혈장, 혈청, 뇌척수액, 흉막액, 복수액, 유도 분비액, 타액, 기관지 폐포 세척액, 가래, 눈물, 땀 및 뇨를 포함할 수 있다. 체액 외에, 당해 기술을 또한 대변 시료에 적용할 수 있는데, 대변 시료는 결장직장암으로부터의 종양 DNA를 함유하는 것으로 밝혀졌기 때문이다(참조: Berger BM, Ahlquist DA. Pathology 2012; 44: 80-88).

[0041] II. 일반적인 스크리닝 방법

[0042] 도 1은 본 발명의 구현예에 따라 피검자에서 암 또는 악성 변화를 검출하기 위한 방법 (100)의 흐름도이다. 구현예는 피검자로부터의 생물학적 시료 속에서 무세포 DNA를 분석하여 종양으로부터 생성될 수 있는 무세포 DNA내 변이를 검출할 수 있다. 분석은 피검자의 구성적 게놈을 사용하여 건강한 세포의 부분인 다형성에 대해 계수할 수 있으며, 서열분석 오차에 대해 계수할 수 있다. 방법 (100) 및 본원에 기술된 다른 어떠한 방법은 하나 이상의 프로세서를 포함하는 컴퓨터 시스템을 사용하여 전체적으로 또는 부분적으로 수행할 수 있다.

[0043] 단계 (110)에서, 피검자의 구성적 게놈이 수득된다. 구성적 게놈(CG)은 시험한 피검자의 구성적 DNA로부터 측정될 수 있다. 각종 구현예에서, CG는 예를 들면, 세포-유리된 DNA를 포함하는 시료로부터의 세포 속에 존재할 수 있는, 구성적 DNA의 서열 판독물을 분석함으로써 활성적으로 측정할 수 있거나 기억장치로부터 판독할 수 있다. 예를 들어, 비-혈액확장 악성종양이 예측되는 경우, 혈액 세포를 분석하여 피검자의 구성적 DNA를 측정할 수 있다.

[0044] 각종 시행에서, 구성적 DNA의 분석은 거대한 평행 서열분석, 배열 계 하이브리드화(array-based hybridization), 프로브 계 용액내 하이브리드화, 연결 계 검정, 프라이머 연장 반응 검정, 및 질량 분석법을 사용하여 수행할 수 있다. 하나의 구현예에서, CG는 피검자의 생애 중 한 시점, 예를 들면 출생시 또는 심지어 태아기(이는 태아 세포를 사용하거나 세포-유리된 DNA 단편을 사용하여 수행될 수 있다, 참조: 미국 공개공보 제2011/0105353호)에서 측정된 후 피검자의 생애의 다른 시기에 체액 또는 다른 시료가 수득된 시기를 언급할 수 있다. 따라서, CG는 컴퓨터 기억장치로부터 단순하게 판독될 수 있다. 구성적 게놈은, 구성적 게놈이 기준 게놈과는 상이한 경우 유전자좌의 목록으로서 판독될 수 있다.

[0045] 단계 (120)에서, 하나 이상의 서열 태그가 피검자의 생물학적 시료 속의 다수의 DNA 단편 각각에 대해 수용되며, 여기서 생물학적 시료는 무세포 DNA를 포함한다. 하나의 구현예에서, 하나 이상의 서열 태그는 생물학적 시료중 DNA 단편의 무작위 서열분석으로부터 생성된다. 하나 이상의 서열 태그는, 쌍을 이룬 말단 서열분석을 수행하는 경우 수득될 수 있다. 하나의 태그는 DNA 단편의 각각의 말단에 상응할 수 있다.

[0046] 시료(예를 들면, 혈장, 혈청 또는 다른 체액)속에서 무세포 DNA를 분석하여 유전적 변이에 대해 조사할 수 있다. 무세포 DNA는 구성적 DNA를 분석하는데 사용된 것과 동일한 분석 플랫폼을 사용하여 분석할 수 있다. 달리, 상이한 분석 플랫폼을 사용할 수 있다. 예를 들면, 무세포 DNA 시료를 거대한 평행 서열분석을 사용하여 서열분석하거나 게놈의 부분을 포획하거나 거대한 평행 서열분석 전에 농축시킬 수 있다. 농축을 사용하는 경우, 예를 들면, 게놈의 선택된 부분의 용액 상 또는 고체 상 포획을 사용할 수 있다. 이후에, 거대한 평

행 서열분석을 포획된 DNA 상에서 수행할 수 있다.

- [0047] 단계 (130)에서, 서열 태그에 대한 계놈 위치를 결정한다. 하나의 양태에서, 서열 태그는 기준 계놈에 대해 지정되며, 이는 한 명 이상의 다른 피검자로부터 취득된다. 또 다른 구현예에서, 계놈 서열 태그는 시험한 피검자의 구성적 계놈에 정렬된다. 당해 정렬은 예를 들면, 기본 국제 정렬 조사 도구(Basic Local Alignment Search Tool: BLAST)를 사용하여 당해 분야의 숙련자에게 공지된 기술을 사용하여 수행할 수 있다.
- [0048] 단계 (140)에서, 유전자좌의 제1 구성원이 측정되며, 여기서 적어도 N 서열 태그는 구성적 계놈(CG)과 비교하여 서열 변이를 갖는다. N은 2와 같거나 이보다 더 크다. 하기에 보다 상세히 논의한 바와 같이, 세포 내에서 무작위로 발생하는 서열분석 오차 및 또한 체세포 돌연변이(예를 들면, 세포 분열에 기인함)은 2, 3, 4, 5 이상의 N을 가짐으로써 제거할 수 있다. 하나 이상의 규정된 범주를 만족하는 유전자좌는 돌연변이(변이체) 또는 돌연변이 유전자좌(변이체 유전자좌)로서 확인될 수 있는 반면, 하나 이상의 범주(예를 들면, 바로 하나의 변이체 서열 태그)를 만족시키지 않는 변이체를 갖는 유전자좌는 잠재적인 또는 추정된 돌연변이로 언급된다. 서열 변이체는 바로 하나의 뉴클레오타이드 또는 다수의 뉴클레오타이드일 수 있다.
- [0049] N은 절대값과는 반대되는 것으로서, 유전자좌에 대한 총 태그의 퍼센트로 측정될 수 있다. 예를 들면, 변이체 유전자좌는, 변이체 판독물로부터 부여된 총량 DNA의 분획 농도가 10%(또는 일부 다른 퍼센트) 이상인 것으로 측정되는 경우 확인될 수 있다. 다시 말해서, 유전자 좌리가 200개 서열 판독물에 의해 포함되는 경우, 변이체 대립형질을 나타내는 적어도 10개 서열 판독물의 기준이 돌연변이로서 변이체를 정의하는데 요구될 수 있다. 변이체 대립형질의 10개 서열 판독물 및 야생형 대립형질의 190개 판독물은 10%의 총량 DNA의 분획 농도 ($2 \times 10 / (10 + 190)$)를 제공할 수 있다.
- [0050] 하나의 구현예에서, 서열 태그(총체적으로 시료 계놈으로서 언급됨)를 CG와 직접 비교하여 변이체를 측정할 수 있다. 다른 양태에서, 시료 계놈(SG)을 참조 서열(RG)을 통해 CG와 비교하여 변이체를 측정할 수 있다. 예를 들면, CG 및 SG 둘 다를 RG와 비교하여 변이체를 나타내는 유전자좌의 각각의 수(예를 들면, 세트)를 측정한 후 차이를 취하여 유전자좌의 첫 번째 수를 취득할 수 있다. 첫 번째 수는 수로서 단순히 취득될 수 있거나 유전자좌의 특이적인 세트에 상응할 수 있으며, 이는, 이후에 추가로 분석되어 첫 번째 유전자좌에서 서열 태그로부터 매개변수를 결정할 수 있다.
- [0051] 하나의 시행에서, 구성적 DNA 및 혈장 DNA의 서열분석 결과를 비교하여 단일 뉴클레오타이드 돌연변이가 혈장 DNA 속에 존재하는지를 측정한다. 구성적 DNA가 동형접합성인 영역을 분석할 수 있다. 나열 목적을 위해, 특수 유전자좌의 유전형을 추정하는 것은 구성적 DNA내 동형접합성이며 AA이다. 이후 혈장 속에서, A 이외의 대립형질의 존재는 특수 유전자좌에서 단일의 뉴클레오타이드 돌연변이(SNM)의 잠재적인 존재를 나타낼 수 있다. SNM의 잠재적인 존재를 나타내는 유전자좌는 단계 (140)에서 유전자좌의 첫 번째 수를 형성할 수 있다.
- [0052] 하나의 구현예에서, 특수 암 유형 또는 특히 집단의 서브세트에서 특히 돌연변이인 경향이 있는 것으로 알려진 계놈의 부분을 표적화하는 것이 유용할 수 있다. 후자의 국면과 관련하여, 구현예는 특수한 집단 그룹에서 특히 우세한 돌연변이, 예를 들면, B형 간염 바이러스(간암의 경우) 또는 사람 파필로마바이러스(경부암의 경우)의 매개인자이거나 체세포 돌연변이에 대해 유전적 소인을 갖는 피검자 또는 DNA 미스매치 보수 유전자내 배선 돌연변이를 지닌 피검자에서 특히 일반적인 돌연변이의 유형을 찾을 수 있다. 당해 기술은 또한 BRCA1 또는 BRCA2 돌연변이를 지닌 피검자에서 난소 및 유방암내 돌연변이에 대해 스크리닝하는데 유용할 수 있다. 당해 기술은 APC 돌연변이를 지닌 피검자에서 직장결장암내 돌연변이에 대해 스크리닝하는데 유사하게 유용할 수 있다.
- [0053] 단계 (150)에서, 매개변수는 제1 유전자좌에서 서열 변이를 갖는 서열 태그의 수를 기초로 측정한다. 하나의 예에서, 매개변수는 유전자좌의 첫 번째 수이며, 여기서 적어도 N개 DNA 단편은 구성적 계놈에 대해 유전자좌에서 서열 변이를 갖는다. 따라서, 계수를 단순히 사용하여 유전자좌가 첫 번째 수에 포함되기 전에 확인된 특수 변이체의 N개 카피보다 더 많은지를 보증할 수 있다. 다른 구현예에서, 매개변수는 제1 유전자좌에서 구성적 계놈에 대해 서열 변이를 갖는 서열 태그의 전체 수일 수 있거나 이를 포함할 수 있다.
- [0054] 단계 (160)에서, 피검자에 대한 매개변수를 피검자에서 암의 수준의 분류를 결정하기 위한 역치 값(예를 들면, 한 명 이상의 다른 피검자로부터 기원함)과 비교한다. 암 수준의 예는, 피검자가 암을 지니고 있는지 또는 악성 상태인지, 또는 암으로 진행될 확률이 증가되어 있는지를 포함한다. 하나의 구현예에서, 역치 값은 피검자로부터 이미 취득된 시료로부터 결정될 수 있다.
- [0055] 다른 구현예에서, 한 명 이상의 다른 피검자는 암을 지니지 않거나 암의 위험이 낮은 것으로 측정될 수 있다. 따라서, 역치 값은 정상 값, 정상 범위일 수 있거나, 정상 값 또는 범위로 통계를 유의적인 편차를 나

타낼 수 있다. 예를 들어, 암을 지니지 않거나 암 위험이 낮은 피검자의 혈장 속에서 검출가능한, 특수 피검자의 CG에 대한 돌연변이의 수를 일반 범위로 사용하여 시험한 피검자에서 검출된 돌연변이의 수가 일반적인지를 측정할 수 있다. 다른 구현예에서, 다른 피검자는 암을 지닌 것으로 공지될 수 있으므로, 유사한 수의 돌연변이는 암을 나타낼 수 있다.

[0056] 하나의 시행에서, 다른 피검자를 선택하여 시험 피검자의 임상 특성, 예를 들면, 성별, 연령, 식이, 흡연 습관, 약물 사용력(drug history), 이전 약물, 가족력, 선택된 게놈 유전자좌의 유전형, 바이러스 감염(예를 들면, B 또는 C형 간염 바이러스 또는 사람 파필로마바이러스 또는 사람 면역결핍성 바이러스 또는 엡슈타인-바르 바이러스 감염) 또는 다른 감염성 체제[예를 들면, 세균(예를 들면, 헬리코박터 필로리(*Helicobacter pylori*) 및 기생충(예를 들면, 클로노르키스 시넨시스(*Clonorchis sinensis*) 등]에 의한 감염과 일치하는 임상 특성을 가지도록 할 수 있다. 예를 들면, B 또는 C형 간염 바이러스의 매개체인 피검자는 간세포 암종으로 발달할 위험성이 증가되어 있다. 따라서, B 또는 C형 간염의 매개체로서 돌연변이의 유사한 수 또는 양식을 갖는 시험 피검자는 간세포 암종으로 발달할 위험성이 증가된 것으로 고려될 수 있다. 다른 한편, 다른 간염 환자보다 더 많은 돌연변이를 나타내는 B 또는 C형 간염 환자는, 적절한 기본선(즉, 다른 간염 환자에 비해)이 사용되므로, 암의 수준의 보다 높은 분류를 갖는 것으로 적절히 확인될 수 있다. 유사하게, 사람 파필로마바이러스 감염의 매개체인 피검자는 경부 암 및 두경부 암에 대한 위험성이 증가되어 있다. 엡슈타인-바르 바이러스에 의한 감염은 비인두 암종, 위암, 호지킨 림프종(Hodgkin's lymphoma) 및 비-호지킨 림프종(non-Hodgkin's lymphoma)과 관련되어 왔다. 헬리코박터 필로리(*Helicobacter pylori*)에 의한 감염은 위암과 관련되어 왔다. 클로노르키스 시넨시스(*Clonorchis sinensis*)에 의한 감염은 담관암종과 관련되어 왔다.

[0057] 상이한 시점에서 돌연변이의 수의 변화의 모니터링은 암의 진행 및 치료 반응을 모니터링하는데 사용될 수 있다. 이러한 모니터링을 또한 사용하여, 피검자가 암으로 발달할 위험성에 있어서 악성 변화 또는 상태의 진행을 문서화할 수 있다.

[0058] 변이를 나타내는 서열 태그의 양을 또한 사용하여 모니터링할 수 있다. 예를 들면, 유전자좌에서 변이체 판독의 분획 농도를 사용할 수 있다. 하나의 구현예에서, 일련의 모니터링 동안 시료 속에서 종양-관련된 유전적 일탈의 분획 농도에 있어서의 증가는 질병의 진행 또는 임박한 재발을 나타낸다. 유사하게, 일련의 모니터링 동안 시료 속에서 종양-관련된 유전적 일탈의 분획 농도에 있어서의 감소는 치료 및/또는 완화 및/또는 우수한 징후에 대한 반응을 나타낼 수 있다.

[0059] III. 게놈의 측정

[0060] 상기 논의된 각종 게놈을 하기에 보다 상세히 설명한다. 예를 들면, 기준 게놈, 구성적 게놈, 및 시료 게놈을 논의한다.

[0061] A. 기준 게놈

[0062] 기준 게놈(RG)는 피검자의 반수체 또는 이배체 게놈 또는 집단의 컨센서스를 말한다. 기준 게놈은 공지되어 있으므로 새로운 환자로부터의 서열분석 판독물을 비교하는데 사용할 수 있다. 환자의 시료로부터 서열 판독물을 정렬하여 비교함으로써 RG로부터 판독물에 있어서의 변이를 확인할 수 있다. 반수체 게놈의 경우, 각각의 유전자 자리에 단지 하나의 뉴클레오타이드가 존재하고 이에 따라 각각의 유전자좌는 반접합성인 것으로 고려될 수 있다. 이배체 게놈의 경우, 이형접합성 유전자좌는 2가지 대립형질을 갖는 이러한 유전자좌로 확인될 수 있으며, 여기서 대립형질은 유전자좌에 대한 정렬용 조화를 허용할 수 있다.

[0063] 기준 게놈은 피검자의 집단 중에서 동일할 수 있다. 이러한 동일한 기준 게놈은, 건강한 피검자가 환자(예를 들면, 암을 지니거나 지니지 않은)를 분류하는데 사용하기 위한 적절한 역치를 결정하는데 사용할 수 있다. 그러나, 상이한 기준 게놈은 상이한 집단, 예를 들면 상이한 민족성 또는 심지어 상이한 가계에 대해 사용할 수 있다.

[0064] B. 구성적 게놈

[0065] 피검자(예를 들면, 사람 또는 다른 이배체 유기체)에 대한 구성적 게놈(CG)은 피검자의 이배체 게놈을 말한다. CG는 이형접합성 유전자좌를 규정할 수 있으며, 여기서 제1 대립형질은 제1 반수체 유형으로부터 기원하고 상이한 제2의 대립형질은 제2의 반수체 유형으로부터 기원한다. 2개의 이형접합성 유전자좌를 포함하는 2개의 반수체 유형의 구조, 즉, 하나의 이형접합성 유전자좌에서 어느 대립형질이 공지될 필요가 없는 다른 이형접합성 유전자좌의 대립형질과 동일한 반수체 유형 위에 존재하는지는 알려질 필요가 없다. 각각의 이형접합성 유전자좌

에서 2가지 대립형질의 존재만으로도 충분할 수 있다.

[0066] CG는 다형체로 인하여 RG와는 상이할 수 있다. 예를 들면, RG 상에서 유전자좌는 T에 대해 동형접합성일 수 있지만, CG는 T/A에 대해 이형접합성이다. 따라서, CG는 당해 유전자좌에서 변이를 나타낼 수 있다. CG는 또한 유전된 돌연변이(예를 들면, 가계 내에서 발생하는) 또는 드 노보 돌연변이(de novo mutation)(이는 태아에서 발생하지만, 이의 부모에서는 존재하지 않는다)로 인하여 RG와는 상이할 수 있다. 유전된 돌연변이는 전형적으로 '배선 돌연변이'로 불린다. 이러한 돌연변이중의 일부는 암에 대한 성향, 예를 들면, 가계 내에서 발생하는 BRCA1 돌연변이와 관련된다. 이러한 돌연변이는 생물의 일생 동안 세포 분열로 인해 발생할 수 있는 '체세포 돌연변이'와는 상이하며 세포 및 이의 후대세포를 암이 되도록 하는 방식으로 압박할 수 있다.

[0067] CG를 결정하는 목표는 체세포 돌연변이를 확인하기 위하여 시료 게놈(SG)의 돌연변이로부터 이러한 배선 돌연변이 및 드 노보 돌연변이를 제거하는 것이다. SG에서 체세포 돌연변이의 양을 이후에 사용하여 피검자에서 암의 경향성을 평가할 수 있다. 이러한 체세포 돌연변이는 추가로 여과하여 서열분석 오차를 제거하고, 이러한 체세포 돌연변이는 암과 관련되지 않는 경향이 있으므로, 잠재적으로 드물게 발생하는 체세포 돌연변이(예를 들면, 변이체를 나타내는 단지 하나의 관독물)를 제거할 수 있다.

[0068] 하나의 구현예에서, CG는 세포(연막(buffy coat) DNA)를 사용하여 결정할 수 있다. 그러나, CG는 또한 세포-유리된 DNA(예를 들면, 혈장 또는 혈청)으로부터 또한 측정할 수 있다. 대부분의 세포가 비-악성인 시료 유형, 예를 들면, 건강한 피검자로부터의 연막인 경우, 대부분 또는 컨센서스 게놈은 CG이다. CG의 경우, 각각의 게놈 유전자좌는 시료채취된 조직 내 세포 대부분이 지닌 DNA 서열로 이루어진다. 서열분석 깊이는 구성적 게놈 내 이형접합성 부위를 설명하는데 충분하여야 한다.

[0069] 다른 예로서, 혈장을 구성적 시료로 사용하여 CG를 측정할 수 있다. 예를 들면, 혈장 중 종양 DNA가 50% 미만이고 SNM이 이형접합성 상태인 경우, 예를 들면 돌연변이가 새로운 대립형질의 첨가인 경우, 새로운 대립형질은 25% 미만의 농도를 가질 수 있다. 반면, CG내 SNP의 이형접합성 대립형질의 농도는 대략 50%의 양이어야 한다. 따라서, CG의 체세포 돌연변이와 다형체 사이에 구별이 이루어질 수 있다. 하나의 시행에서, 적합한 컷오프는 혈장, 또는 유의적인 종양 농도를 지닌 다른 혼합물을 사용하는 경우 다형체로부터 체세포 돌연변이를 결정하는 경우 30 내지 40%일 수 있다. 종양 DNA 농도의 측정은, 혈장 속의 종양 DNA가 50% 미만이 되도록 보증하는데 유용할 수 있다. 종양 DNA 농도를 결정하는 예는 본원에 기술되어 있다.

[0070] C. 시료 게놈

[0071] 시료 게놈(SG)은 RG 및 CG의 경우에서와 같이 단순히 반수체 또는 이배체 게놈이 아니다. SG는 시료로부터의 관독물의 수집물이며, CG에 상응하는 구성적 DNA로부터의 관독물, 종양 DNA로부터의 관독물, CG에 대해 무작위 돌연변이를 나타내는 건강한 세포로부터의 관독물(예를 들면, 세포 분열로부터 생성되는 돌연변이로 인한), 및 서열분석 오차를 포함할 수 있다. 각종 매개변수를 사용하여 어떠한 관독물이 SG에 포함되는지를 정확하게 조절할 수 있다. 예를 들어, 적어도 5개 이하의 관독물에서 나타내도록 대립형질을 요구하는 것은 SG 속에 존재하는 서열분석 오차를 감소시킬 수 있으며 또한 무작위 돌연변이로 인한 관독물을 감소시킬 수 있다.

[0072] 예로서, 대상체가 건강한 것으로, 즉, 암을 지니지 않은 것으로 추정한다. 나열 목적을 위해, 1000개 세포로부터의 DNA는 당해 피검자로부터 수득된 혈장 1ml(즉, DNA의 1000개 게놈-등가물)이다. 혈장 DNA는 전형적으로 약 150bp의 DNA 단편으로 이루어진다. 사람 게놈이 3×10^9 bp이므로, 반수체 게놈당 약 2×10^7 개 DNA 단편이 존재할 수 있다. 사람 게놈이 이배체인 경우, 혈장 ml당 약 4×10^7 개 DNA 단편이 존재할 수 있다.

[0073] 수백만 내지 수십억개의 세포가 단위 시간당 혈장 속의 이들의 DNA로부터 방출되고 이들 세포로부터의 단편은 순환 동안 함께 혼합될 수 있으므로, 4×10^7 개의 DNA 단편은 4×10^7 개의 상이한 세포로부터 올 수 있다. 이들 세포는 서로에 대해 최근(먼 것, 예를 들면, 원래의 접합체와 반대되는 것으로서)의 클론 관계를 지니지 않는(즉, 이들이 최근의 조상 세포를 공유하지 않는) 경우, 이는, 이들 단편들 중에서 1회 이상 돌연변이가 관찰되지 않을 것이라는 것과 통계적으로 유사하다.

[0074] 다른 한편, 혈장 DNA의 ml당 1000개 게놈-등가물 중에서, 최근의 조상 세포를 공유하는(즉, 이들은 각각 서로 클론적으로 관련되어 있다) 특정 퍼센트의 세포가 존재하는 경우, 당해 클론으로부터의 돌연변이를 찾아서 혈장 DNA 내에 우선적으로 나타낼 수 있다(예를 들면, 혈장 속에서 클론 돌연변이 프로파일을 나타냄). 이러한 클론적으로 관련된 세포는 암 세포, 또는 이들의 방식으로 암이 되지만 아직 존재하지는 않는(즉, 전-신생물성) 세포일 수 있다. 따라서, 돌연변이가 1회 이상 까지 나타나도록 요구하는 것은 시료 속에서 확인된 "돌연변이"내

이러한 천연 변이를 제거할 수 있으며 이는, 암 세포 또는 전-신생물 세포와 관련된 보다 많은 돌연변이를 제거함으로써 검출, 특히 암 또는 전암성 상태의 조기 검출을 허용한다.

[0075] 하나의 근사치로서, 평균적으로, 하나의 돌연변이가 매 세포 분열 후 게놈 내에 축적될 것으로 설명되어 왔다. 앞서의 연구는, 혈장 DNA의 대부분이 조혈 세포로부터 기원함을 입증하여 왔다(참조: Lui YY 등 Clin Chem 2002: 48: 421-427). 조혈 줄기 세포는 매 25 내지 50주마다 1회 복제함이 추정되어 왔다(참조: Catlin SN, 등 Blood 2011; 117: 4460-4466). 따라서, 단순 예측으로서, 건강한 40세 피검자는 조혈 줄기 세포당 일부 40 내지 80개의 돌연변이를 축적할 수 있다.

[0076] 이러한 개인의 혈장 속에 ml당 1000개의 게놈-등가물이 존재하고 이들 세포 각각이 상이한 조혈 줄기 세포로부터 기원하는 경우, 40,000 내지 80,000개의 돌연변이가 4×10^{10} 개의 DNA 단편(즉, 게놈당 4×10^7 개 DNA 단편, 및 혈장 ml당 1000개의 게놈-등가물) 중에서 예측될 수 있다. 그러나, 각각의 돌연변이가 1회 관측될 수 있으므로, 각각의 돌연변이가 여전히 검출 한계 미만으로 존재할 수 있고(예를 들면, 컷오프 값 N이 1보다 큰 경우), 이에 따라 이들 돌연변이를 여과함으로써, 분석이 음성 상태에서부터 생성되는 경향성이 보다 더 있는 돌연변이에 초점을 맞추도록 할 수 있다. 컷오프 값은 1 초과와 특정 값(정수 또는 비-정수)일 수 있으며, 상이한 유전자좌 및 영역에 대해 역학적일 수 있다. 종양 DNA의 서열분석 깊이 및 분획 농도는 암 세포 또는 전-신생물 세포로부터 돌연변이(예를 들면, 검출가능한 돌연변이의 퍼센트)를 검출하는 민감성에 영향을 미칠 수 있다.

[0077] IV. CG에 대해 직접적인 SG의 비교

[0078] 일부 구현에는, CG가 동형접합성이지만, SG내 소수 종(즉, 종양 DNA)가 이형접합성인 뉴클레오타이드 위치를 확인할 수 있다. 높은 깊이에서의 위치(예를 들면, 50-배 이상 포함)를 서열분석하는 경우, 건강한 세포와 암 세포의 DNA 혼합물 속에서 당해 위치에서 1개 또는 2가지 대립형질이 존재하는지를 검출할 수 있다. 2가지 대립형질이 검출되는 경우, (1) CG는 이형접합성이거나 (2) CG는 동형접합성이지만 SG는 이형접합성이다. 주요 및 약간의 대립형질의 상대적인 수를 고찰함으로써 이들 2개의 시나리오를 차별화할 수 있다. 전자의 시나리오에서, 2가지 대립형질은 유사한 다수의 수를 가질 수 있지만; 후자의 시나리오의 경우, 이들의 카운트수에 있어서 큰 차이가 있을 수 있다. 시험 시료로부터의 판독물의 상대적인 대립형질 수의 이러한 비교는 서열 태그를 구성적 게놈과 비교하기 위한 하나의 구현예이다. 방법 (100)의 첫 번째 유전자좌는, 대립형질의 수가 상한치(CG에서 다형성에 상응하는 역치) 미만이고 하한치(암 상태와 관련되지 않는 충분히 낮은 비율로 발생하는 오차 및 체세포 돌연변이에 상응하는 역치) 초과인 유전자좌로 측정될 수 있다. 따라서, 구성적 게놈 및 제1 유전자좌는 동시에 측정될 수 있다.

[0079] 다른 구현예에서, 돌연변이를 확인하기 위한 공정은, 우선 CG를 측정 한 후, CG에 대해 상대적인 충분한 수의 돌연변이를 갖는 유전자좌를 측정할 수 있다. CG는 시험 시료와는 상이한 구성적 시료로부터 결정될 수 있다.

[0080] 도 2는 본 발명의 구현예에 따라서 시료 게놈(SG)을 구성적 게놈(CG)과 직접적으로 비교하는 방법 (200)의 흐름도를 나타낸다. 블록 (210)에서, 피검자의 구성적 게놈이 획득된다. 구성적 게놈은 예를 들면, 적시에 이미 취한 시료로부터 또는 방법 (200)이 시행되기 직전에 획득되고 분석된 구성적 시료로부터 획득될 수 있다.

[0081] 블록 (220)에서, 하나 이상의 서열 태그는 피검자의 생물학적 시료 중의 다수의 DNA 단편 각각에 대해 수용된다. 서열분석은 본원에 언급된 바와 같이, 각종 기술을 사용하여 수행할 수 있다. 서열 태그는, 단편의 어느 서열이 존재하는 것으로 여겨지는지의 척도이다. 그러나, 서열 태그의 하나 이상의 염기는 오차 내에 존재할 수 있다.

[0082] 블록 (230)에서, 서열 태그의 적어도 하나의 부위를 구성적 게놈에 정렬한다. 정렬은 각종 유전자좌에서 이형접합성인 CG를 계수할 수 있다. 정렬은, 변이체가 검출될 수 있도록 하는 정확한 조화를 필요로 하지 않을 수 있다.

[0083] 블록 (240)에서, 구성적 게놈에 대해 상대적인 유전자좌에서 서열 변이를 갖는 서열 태그가 확인된다. 서열 태그는 하나 이상의 변이체를 가질 수 있는 것이 가능하다. 각각의 유전자좌 및 각각의 서열 태그에 대한 변이체를 추적할 수 있다. 변이체는 CG내 존재하지 않는 특정 대립형질일 수 있다. 예를 들면, CG는 A/T에 대해 이형접합성일 수 있으며 변이체는 G 또는 C일 수 있다.

[0084] 블록 (250)에서, 변이체를 가진 각각의 유전자좌의 경우, 컴퓨터 시스템은 유전자좌를 정렬하는 서열 태그의 각각의 첫 번째 수를 계수할 수 있으며 유전자좌에서 서열 변이를 가질 수 있다. 따라서, 각각의 유전자좌는 유

전자좌에서 관찰된 변이체들의 수와 관련된 수를 가질 수 있다. 전형적으로, 보다 적은 변이체가 예를 들면, 50% 미만인 종양 DNA 농도로 인하여, CG에 상응하는 서열 태그와 비교된 유전자좌에서 관찰될 것이다. 그러나, 일부 시료는 50% 초과인 종양 DNA의 농도를 가질 수 있다.

[0085] 블록 (260)에서, 매개변수는 각각의 첫 번째 수를 기초로 측정한다. 하나의 구현예에서, 각각의 수가 컷오프 값보다 큰 경우(예를 들면, 2 초과), 각각의 수를 매개변수이거나 매개변수를 결정하는데 사용되는 합에 가할 수 있다. 다른 구현예에서, 컷오프 값보다 큰 각각의 수를 갖는 유전자좌의 수를 매개변수로서 사용한다.

[0086] 블록 (270)에서, 매개변수를 역치 값과 비교하여 암의 수준을 분류한다. 위에 기술된 바와 같이, 역치 값은 다른 피검자로부터의 시료의 분석으로부터 측정할 수 있다. 이들 다른 피검자의 암 상태 또는 건강 상태에 따라, 분류를 측정할 수 있다. 예를 들면, 다른 피검자가 4기 암을 가진 경우, 이후에 현재의 매개변수가 다른 피검자로부터 수득된 매개변수의 값과 밀접한 경우(예를 들면, 규정된 범위내), 현재의 피검자는 4기 암을 가진 것으로 분류될 수 있다. 그러나, 매개변수가 역치를 초과하는 경우(즉, 매개변수가 정의된 방법에 따라, 보다 더 크거나 작은 경우), 분류는 4기 미만으로 정의될 수 있다. 다른 피검자가 암을 지니지 않은 경우 유사한 분석이 이루어질 수 있다.

[0087] 다중 역치를 사용하여 분류를 측정할 수 있으며, 여기서 각각의 역치는 상이한 세트의 피검자로부터 측정한다. 피검자의 각각의 세트는 일반적인 암 수준을 가질 수 있다. 따라서, 현재의 매개변수를 피검자의 각각의 세트에 대한 값과 비교할 수 있으며, 이는 세트 중 하나에 대한 조화를 제공할 수 있거나 범위를 제공할 수 있다. 예를 들면, 매개변수는 전암성 또는 2기인 피검자에 대해 수득된 매개변수와 거의 동일할 수 있다. 다른 예로서, 현재의 매개변수는 암의 몇가지 상이한 수준과 가능하게는 조화될 수 있는 범위에 속할 수 있다. 따라서, 분류는 암의 하나 이상의 수준을 포함할 수 있다.

[0088] V. 기준 게놈의 사용

[0089] 구성적 DNA 및 생물학적 시료로부터의 DNA 둘 다의 게놈 서열을 사람 기준 게놈과 비교할 수 있다. 기준 게놈과 비교하여 구성적 DNA보다 혈장 시료 속에서 보다 많은 변화가 존재하는 경우, 암 확률이 더 높다. 하나의 구현예에서, 기준 게놈 내 동형접합성 유전자좌를 연구한다. 구성적 DNA 및 생물학적 시료로부터의 DNA 둘 다에 있어서 이형접합성 유전자좌의 양을 비교한다. 생물학적 시료의 DNA로부터 검출된 이형접합성 부위의 양이 구성적 DNA의 것을 초과하는 경우, 암 확률이 보다 더 높다.

[0090] 분석은 또한 CG에 있어서 동형접합성 유전자좌로 한정될 수 있다. SNM은 또한 이형접합성 유전자좌에 대해 정의될 수 있지만, 이는 일반적으로 제3의 변이체의 생성을 필요로 할 수 있다. 다시 말해서, 이형접합성 유전자좌가 A/T인 경우, 신규 변이체는 C 또는 G일 수 있다. 동형접합성 유전자좌에 대한 SNM을 확인하는 것은 일반적으로 더 쉽다.

[0091] 구성적 DNA와 비교하여 생물학적 시료 DNA중 이형접합성 유전자좌의 양에 있어서의 증가 정도는, 건강한 피검자에서 관찰된 변화율과 비교하는 경우 암 또는 악성 상태의 제안일 수 있다. 예를 들어, 이러한 부위에 있어서의 증가 정도가 건강한 피검자에서 관찰된 것을 특정 역치까지 초과하는 경우, 당해 데이터를 암 또는 악성 상태의 제안인 것으로 고려할 수 있다. 하나의 구현예에서, 암이 없는 피검자에서 돌연변이의 분포가 확인되며 역치는 특정 수의 표준 편차(예를 들면, 2 또는 3의 표준 편차)로 고려될 수 있다.

[0092] 하나의 구현예는, 유전자좌를 계수하기 전에 유전자좌에서 변이체의 적어도 규정된 수를 필요로 할 수 있다. 다른 구현예는 변화를 일단 관찰하는 것을 기초로 하여 데이터에 대해 심지어 시험을 제공한다. 예를 들어, 혈장 속에서 관찰된 변이의 총 수(오차 + 실제 돌연변이 또는 다형체)가 구성적 DNA에서의 것보다 유의적으로 더 높은 경우, 암의 증거가 존재한다.

[0093] 도 3은 본 발명의 구현예에 따른 기준 게놈(RG)을 사용하여 시료 게놈(SG)을 구성적 게놈(CG)과 비교하는 방법 (300)의 흐름도를 나타낸다. 방법 (300)은, RG가 이미 수득된 것으로 추정하며, 생물학적 시료에 대한 서열 태그가 이미 수용된 것으로 추정한다.

[0094] 블록 (310)에서, 서열 태그의 적어도 한 부위를 기준 게놈에 대해 정렬한다. 당해 정렬은 변이체로서의 미스매치(mismatch)가 검출되도록 할 수 있다. 기준 게놈은 피검자와 유사한 집단으로부터 존재할 수 있다. 정렬된 서열 태그는 시료 게놈(SG)을 효과적으로 포함한다.

[0095] 블록 (320)에서, 잠재적인 변이체의 첫 번째 수(A), 예를 들면, 단일 뉴클레오타이드 돌연변이(SNM)가 확인된다. 잠재적인 SNM은, SG의 서열 태그가 RG와는 상이한 뉴클레오타이드를 나타내는 유전자좌이다. 다른

기준, 예를 들면, 변이를 나타내는 서열 태그의 수는 컷오프 값보다 더 커야하며 유전자좌가 RG 내에서 동형접합성인지의 여부를 사용할 수 있다. 구체적인 유전자좌가 확인되고 저장장치 내에서 유전자좌를 저장함으로써 추적되는 경우 잠재적인 SNM의 세트를 세트 A로서 나타낼 수 있다. 구체적인 유전자좌가 측정될 수 있거나 단순히 다수의 이러한 SNM을 측정할 수 있다.

[0096] 블록 (330)에서, 구성적 게놈은 구성적 시료로부터의 DNA 단편을 서열분석함으로써 수득된 서열 태그를 기준 게놈에 정렬함으로써 측정한다. 당해 단계는 앞서 어느 시기에서도 수득된 구성적 시료를 사용함으로써 수행될 수 있다. CG는 단순히 기억장치로부터 판독될 수 있으며, 여기서 정렬은 이미 수행되었다. 하나의 구현예에서, 구성적 시료는 혈액 세포일 수 있다.

[0097] 블록 (340)에서, CG의 정렬된 서열 태그가 기준 게놈과 비교하여 유전자좌에서 변이체(예를 들면, SNM)를 갖는 유전자좌의 두 번째 수(B)가 확인된다. 유전자좌의 세트가 구체적으로 추적되는 경우, B는 이 수와 대치되는 것으로서, 세트를 나타낼 수 있다.

[0098] 블록 (350)에서, 세트 B는 세트 A로부터 감하여 시료 게놈 속에 존재하지만 CG에는 존재하지 않는 변이체(SNM)를 확인한다. 하나의 구현예에서, SNM의 세트는, CG가 동형접합성인 뉴클레오타이드 위치에 한정될 수 있다. 이러한 필터링을 달성하기 위하여, CG가 동형접합성인 구체적인 유전자좌를 세트 C에서 확인할 수 있다. 다른 구현예에서, CG가 유전자좌에서 동형접합성이 아닌 경우, 유전자좌는 첫 번째 수 A 또는 두 번째 수 B에서 계수되지 않는다. 다른 구현예에서, 어떠한 공지된 다형체(예를 들면, SNP 데이터베이스 내에서 이의 존재로 인한 함)는 필터링하여 제거될 수 있다.

[0099] 하나의 구현예에서, 블록 (350)에서 추출은 단순히 수의 차감일 수 있으므로, 구체적인 잠재적 SNM은 제거되지 않으나, 단순히 값이 감해진다. 다른 구현예에서, 차감은 세트 A와 세트 B 사이에서 차이를 취하여(예를 들면, 세트 B가 세트 A의 서브세트인 경우) 세트 B가 아닌 구체적인 SNM을 확인한다. 논리 값에서, 이는 $[A \text{ AND NOT}(B)]$ 로 나타낼 수 있다. 확인된 변이체의 생성된 세트는 C로 표지될 수 있다. 매개변수는 수 C로 측정될 수 있거나 세트 C로부터 측정될 수 있다.

[0100] 일부 구현예에서, 돌연변이의 특성이 고려될 수 있으며 상이한 가중치가 상이한 돌연변이 부류에 기여될 수 있다. 예를 들면, 암과 일반적으로 관련된 돌연변이는 보다 높은 가중치(또한 유전자좌의 상대적인 가중치를 참조하는 경우 중요 값으로 불림)에 기여할 수 있다. 이러한 돌연변이는 종양-관련된 돌연변이의 데이터베이스, 예를 들면, 암에서 체세포 돌연변이의 카탈로그(Catalogue of Somatic Mutations in Cancer: COSMIC)(www.sanger.ac.uk/genetics/CGP/cosmic/)에서 찾을 수 있다. 다른 예로서, 비슷하지 않은 변화와 관련된 돌연변이는 보다 높은 가중치에 기여할 수 있다.

[0101] 따라서, 첫 번째 수 A는 가중된 합으로 측정될 수 있으며, 여기서 하나의 유전자좌에서 변이체를 나타내는 태그의 계수(count)는 다른 유전자좌에서 태그의 계수보다 상이한 가중치를 가질 수 있다. 첫 번째 수 A는 이러한 가중된 합을 반영할 수 있다. 유사한 계산을 B에서 수행할 수 있으므로, 수 C 및 매개변수는 이러한 가중치를 반영할 수 있다. 다른 구현예에서, 가중치는, 구체적인 유전자좌의 세트 C가 측정되는 경우에 대해 계수된다. 예를 들어, 가중된 합이 세트 C의 유전자좌에 대한 총 수에 대해 측정될 수 있다. 이러한 가중치는 본원에 기술된 다른 방법의 경우 사용될 수 있다.

[0102] 따라서, 암의 수준의 분류를 결정하기 위해 역치와 비교되는 매개변수는 RG와 비교하여 SG 및 CG에 대한 변이를 나타내는 유전자좌의 수일 수 있다. 다른 구현예에서, 변이를 나타내는 DNA 단편의 총 수(서열 태그를 통해 계수된 것으로서)가 계수될 수 있다. 다른 구현예에서, 이러한 수들을 다른 식에서 사용하여 매개변수를 수득할 수 있다.

[0103] 하나의 구현예에서, 각각의 유전자좌에서 변이체의 농도는 매개변수일 수 있으며 역치와 비교될 수 있다. 당해 역치는, 유전자좌가 잠재적인 변이체 유전자좌(변이체를 나타내는 판독물의 구체적인 수의 컷오프 외에)인지를 측정한 후 유전자좌를 계수하는데 사용될 수 있다. 농도는 또한 SNM의 합에서 가중 인자로서 사용될 수 있다.

[0104] VI. 컷오프 값을 사용한 거짓 양성 감소

[0105] 위에서 언급한 바와 같이, 단일 뉴클레오타이드 돌연변이를 다수의 세포-유리된 DNA 단편(예를 들면, 혈장 속에서 순환하는 DNA) 속에서 거대한 게놈 영역(예를 들면, 전체 게놈)에 대해 또는 다수의 게놈 영역에 대해 조사하여 당해 시도의 민감성을 개선시킬 수 있다. 그러나, 서열분석 오차와 같은 분석 오차는 당해 시도의 실행확률, 정밀도 및 특이성에 영향을 미칠 수 있다. 여기서, 본 발명자들은 서열분석 오차의 중요성을 나열하기 위한 예로서 거대한 평행 서열분석 플랫폼을 사용한다. 합성에 의한 일루미나 서열분석 플랫폼(Illumina

sequencing-by-synthesis platform)의 서열분석 오차율은 서열분석된 뉴클레오타이드당 대략 0.1% 내지 0.3%이다(참조: Minoche 등 Genome Biol 2011, 12:R1 12). 연결에 의한 서열 분석 플랫폼(sequencing-by-ligation platform)(예를 들면, Life Technologies SOLiD platform), 이온 토런트(the Ion Torrent)/이온 양성자(Ion Proton), 반도체 서열분석(semiconductor sequencing), Roche 454, 단일 분자 서열분석 플랫폼(예를 들면, Helicos, Pacific Biosciences and nanopore)을 포함하는 어떠한 거대한 평행 서열분석 플랫폼도 사용할 수 있다.

[0106] 간세포 암종에 있어서의 앞서의 연구에서, 전체 암 게놈에 대해 대략 3,000개의 단일 뉴클레오타이드 돌연변이가 존재함이 밝혀졌다(참조: Tao Y 등 2011 Proc Natl Acad Sci USA; 108: 12042-12047). 순환시 전체 DNA의 단지 10% 만이 종양 세포로부터 기원하며 본 발명자가 1배의 반수체 게놈 포함의 평균 서열 분석 깊이를 사용하여 혈장 DNA를 서열분석하는 것으로 추정하면, 본 발명자는 서열분석 오차로 인하여 9백만($3 \times 10^9 \times 0.3\%$)개의 단일 뉴클레오타이드 변이(SNV)에 직면할 수 있다. 그러나, 단일 뉴클레오타이드 돌연변이의 대부분은 2개의 동종 염색체 중 단지 하나에서만 발생하는 것으로 예측된다. 100% 종양 DNA를 지닌 시료의 1-배 반수체 게놈 포함의 서열분석 깊이를 사용하여, 본 발명자는 3,000개 돌연변이의 단지 1/2, 즉, 1,500개 돌연변이만을 검출하는 것으로 예측한다. 본 발명자가 하나의 반수체 게놈 포함에 대한 10%의 종양-기원한 DNA를 함유하는 혈장 시료를 서열분석하는 경우, 본 발명자는 단지 150($1,500 \times 10\%$)개의 암-관련된 단일 뉴클레오타이드 돌연변이만을 검출하는 것으로 예측한다. 따라서, 암-관련된 돌연변이의 검출을 위한 시그널-대-노이즈 비(signal-to-noise ratio)는 60,000 중 1이다. 이러한 매우 낮은 시그널-대-노이즈 비는, 본 발명자가 생물학적 시료(예를 들면, 혈장) 속에서 모든 단일 뉴클레오타이드 변화를 매개변수로서 단순히 사용하는 경우 정상 및 암 경우를 차별화하기 위한 당해 시도를 사용하는 정밀도가 매우 낮을 수 있음을 제안한다.

[0107] 서열분석 기술에 있어서의 진전으로, 서열분석 오차율에 있어서의 지속적인 감소가 있을 수 있는 것으로 예측된다. 또한 하나 이상의 서열분석 플랫폼을 사용하고 교차-플랫폼 서열분석 결과의 비교를 통해 동일한 시료를 분석하여, 서열분석 오차에 의해 영향받는 경향이 있는 관독물을 정확히 찾아낼 수 있다. 다른 시도는 동일한 피검자로부터 상이한 시점에서 취한 2개 시료를 분석하는 것이다. 그러나, 이러한 시도는 시간 소모적이다.

[0108] 하나의 구현예에서, 암 환자의 혈장 속에서 단일의 뉴클레오타이드 돌연변이의 검출시 시그널-대-노이즈 비를 향상시키는 한가지 방법은, 시료 속에 동일한 돌연변이의 다수 발생이 존재하는 경우에만 돌연변이를 계수하는 것이다. 선택된 서열분석 플랫폼에서, 특수한 뉴클레오타이드 치환을 포함하는 서열분석 오차는 보다 일반적이며 시험 피검자 및 대조군 피검자 둘다의 시험 시료 및 구성적 DNA 시료의 서열분석 결과에 영향을 미칠 수 있다. 그러나, 일반적으로, 서열분석 오차는 무작위적으로 발생한다.

[0109] 서열분석 오차를 갖는 기회는, 다수의 DNA 단편 속에서 동일한 뉴클레오타이드 위치에서 동일한 변화를 관찰하는 경우 기하급수적으로 더 낮다. 다른 한편, 시료 속에서 실제 암-관련 돌연변이 변화를 검출하는 기회는 시료 속에서 서열분석 깊이 및 종양 DNA의 분획 농도에 의해 영향받는다. 다수의 DNA 단편에서 돌연변이를 관찰하는 기회는 서열분석 깊이 및 종양 DNA의 분획 농도와 함께 증가할 수 있다. 세포-유리된 종양 DNA를 지닌 시료(예를 들면, 혈장 속)를 사용하는 각종 구현예에서, 분획 농도는 5%, 10%, 20%, 및 30%일 수 있다. 하나의 구현예에서, 분획 농도는 50% 미만이다.

[0110] 도 4는 본 발명의 구현예에 따라 시료 속에 존재하는 것으로서 돌연변이를 분류하기 위한 기준으로서 상이한 수의 발생을 사용하여 정확하게 확인된 암-관련 단일 뉴클레오타이드 돌연변이의 수를 나타내는 표 (400)이다. 동일한 분류 기준을 기준으로 한 서열분석 오차로 인하여 돌연변이를 가진 것으로 거짓 확인된 다수의 뉴클레오타이드 위치의 수를 또한 나타낸다. 서열분석 오차율은 0.1%인 것으로 추측된다(참조: Minoche 등 Genome Bio 2011, 12:R112). 시료 중 종양-기원한 DNA의 분획 농도는 10%인 것으로 추정된다.

[0111] 도 4는, 시료 속에서 종양-기원한 DNA의 분획 농도가 10%인 것으로 추정되는 경우, 혈장 속에서 검출된 암-관련된 돌연변이의 수와 다수의 거짓-양성 요청의 수 사이의 비가 동일한 변화의 수배의 증가가 돌연변이를 정의하기 위해 시료 속에서 동일한 변화가 관찰되면서 기하급수적으로 증가할 수 있음을 나타낸다. 다시 말해서, 암 돌연변이 검출에 대한 민감성 및 특이성 둘 다는 개선될 수 있다. 또한, 암-관련된 돌연변이를 검출하기 위한 민감성은 서열분석 깊이에 의해 영향받는다. 서열분석의 100배 반수체 게놈 포함과 함께, 3,000개의 돌연변이 중 2,205(73.5%)가 시료 속에서 적어도 4개의 DNA 단편내 특수 돌연변이의 발생의 기준을 사용함에 의해서도 검출될 수 있다. 단편의 최소의 수에 대한 다른 값, 예를 들면, 3, 5, 8, 10, 및 10 이상을 사용할 수 있다.

[0112] 도 5는 시료속의 종양 유래 DNA의 분획 농도가 5%인 것으로 추측되는 경우 확인된 돌연변이의 예측된 수 및 거짓-양성 유전자좌의 예측된 수를 나타내는 표 (500)이다. 시료 속에서 종양-기원한 DNA의 보다 낮은 분획 농도

의 사용으로, 보다 높은 서열분석 깊이가 암-관련된 돌연변이를 검출하는 동일한 민감도를 달성하기 위해 요구될 수 있다. 보다 엄격한 기준이 특이성을 유지하는데 또한 요구될 수 있다. 예를 들어, 시료 속에서, 10% 종양 DNA 분획의 상황에서 적어도 4개의 발생의 기준 대신, 적어도 5개의 DNA 단편내 특수한 돌연변이의 발생의 기준을 사용하는 것이 요구될 수 있다. 표 (400) 및 (500)은 제공된 배 포함 및 종양 DNA 농도를 사용하기 위한 컷오프 값에 대한 안내를 제공하며, 이는 본원에 기술된 바와 같이 추정될 수 있거나 측정될 수 있다.

[0113] 단일의 뉴클레오타이드 변화를 검출하는 기준을 1회 이상 사용하여 돌연변이를 정의하는 다른 장점은, 이것이 비-악성 조직 내 단일 뉴클레오타이드 변화로 인하여 거짓 양성 검출을 최소화하는 것으로 예측된다는 것이다. 뉴클레오타이드 변화는 정상 세포의 유사분열 동안 발생할 수 있으므로, 체내에서 각각의 건강한 세포는 다수의 단일 뉴클레오타이드 변화를 지닐 수 있다. 이들 변화는 잠재적으로 거짓 양성 결과를 초래할 수 있다. 그러나, 세포의 변화는, 세포가 사멸하는 경우 혈장/혈청 속에 존재할 수 있다. 상이한 정상 세포가 상이한 돌연변이 세트를 수반하는 것으로 예측되는 반면, 하나의 세포 속에서 발생하는 돌연변이는 혈장/혈청 속에서 다수의 카피로 존재하지 않는 경향이 있다. 이는, 종양 성장이 천연에서 클론성(clonal)이므로 다수의 카피가 혈장/혈청 속에서 관찰되는 것으로 예측되는 종양 세포 내에서의 돌연변이와 대조적이다. 따라서, 클론으로부터의 다수의 세포는 사멸하여 클론을 대표하는 특징적인 돌연변이를 방출한다.

[0114] 하나의 구현예에서, 특이적인 게놈 영역에 대한 표적 농축은 서열분석 전에 수행할 수 있다. 이러한 표적 농축 단계는 수행된 서열분석의 동일한 총량으로 목적인 영역의 서열분석 깊이를 증가시킬 수 있다. 여전히 다른 구현예에서, 비교적 낮은 서열분석 깊이를 사용한 서열분석 라운드(round)를 우선 수행할 수 있다. 이후에, 적어도 하나의 단일 뉴클레오타이드 변화를 나타내는 영역이 보다 높은 배수를 갖는 서열분석의 제2 라운드에 대해 농축될 수 있다. 이후에, 다수의 발생 기준을 적용하여 표적 농축된 서열 분석 결과에 대한 돌연변이를 정의할 수 있다.

[0115] VII. 역학적 컷오프

[0116] 위에 기술한 바와 같이, 변이체(잠재적 돌연변이)를 지지하는 관독물의 수에 대한 컷오프 값 N을 사용하여 유전자좌가 계수될 돌연변이(예를 들면, SNM)로서 자격이 있는지를 측정할 수 있다. 이러한 컷오프를 사용하는 것은 거짓 양성을 감소시킬 수 있다. 하기 논의는 상이한 유전자좌에 대한 컷오프를 선택하기 위한 방법을 제공한다. 다음의 구현예에서, 본 발명자는, 단일의 우세한 암 클론이 존재한다고 추정한다. 유사한 분석을, 상이한 양의 종양 DNA를 혈장 내로 방출하는 암 세포의 다수의 클론을 포함하는 시나리오에 대해 수행할 수 있다.

[0117] A. 혈장 속에서 검출된 암-관련된 돌연변이의 수

[0118] 혈장 속에서 검출가능한 암-관련된 돌연변이의 수는 다수의 매개변수, 예를 들면, 다음의 (1) 내지 (4)에 의해 영향받을 수 있다: (1) 종양 조직 내 돌연변이의 수(N_T) - 종양 조직 내 존재하는 돌연변이의 총 수는 환자의 혈장 속에서 검출가능한 종양-관련된 돌연변이의 최대 수이다; (2) 혈장(f) 속에서 종양-기원한 DNA의 분획 농도 - 혈장 속에서 종양-기원한 DNA의 분획 농도가 높을 수록, 혈장 속에서 종양-관련된 돌연변이를 검출하는 기회가 높을 수 있다; (3) 서열분석 깊이 (D) - 서열분석 깊이는, 서열분석된 영역이 서열 관독물에 의해 포함되는 횟수를 말한다. 예를 들어, 10배의 평균 서열분석 깊이는, 서열분석된 영역 내 각각의 뉴클레오타이드가 평균적으로 10개의 서열 관독물에 의해 포함됨을 의미한다. 암-관련된 돌연변이를 검출하는 기회는, 서열분석 깊이가 증가되는 경우 증가될 수 있다; 및 (4) 잠재적인 암-관련된 돌연변이(r)로서 이를 정의하기 위하여 혈장 속에서 검출되는 뉴클레오타이드 변화의 최소 횟수, 이는 실제 암-관련된 돌연변이로부터 서열분석 오차를 구별하는데 사용된 컷오프 값이다.

[0119] 하나의 실행에서, 푸아송 분포를 사용하여 혈장 속에서 검출된 암-관련 돌연변이의 수를 예측한다. 돌연변이가 서열분석 깊이 D와 함께, 2개의 동종 염색체중 하나 위의 뉴클레오타이드 위치에 존재하는 것으로 추정하여, 혈장 속에 존재하는 예측된 돌연변이 횟수(M_p)는 $M_p = D \times f/2$ 로서 계산한다.

[0120] 특수 돌연변이 부위에서 혈장 속 돌연변이를 검출하는 확률(P_b)은 다음과 같이 계산한다:

$$[0121] P_b = 1 - \sum_{i=0}^{r-1} \text{푸아송}(i, M_p)$$

[0122] 상기 식에서, r(컷오프 값)은, 뉴클레오타이드 변화가 혈장 속에서 관찰되어 이를 잠재적인 종양-관련 돌연변이

로서 정의한 횃수이고; 푸아송(i, M_p)은 M_p 의 평균 수로 발생 i 를 가질 푸아송 분포 확률이다.

[0123] 혈장(NP) 속에서 검출될 것으로 예상된 암-관련된 돌연변이의 총 수는 $N_p = N_T \times P_b$ 로서 계산될 수 있으며, 여기서 N_T 는 종양 조직 속에 존재하는 돌연변이의 수이다. 다음의 그래프는 잠재적인 돌연변이 및 상이한 서열분석 깊이를 나타내기 위한 상이한 발생 기준(r)을 사용하여 혈장 속에서 검출될 것으로 예측된 종양 관련된 돌연변이의 퍼센트를 나타낸다.

[0124] 도 6a는 종양-기원한 DNA의 10% 및 20% 혈장 분획 농도를 사용한 혈장 속에서 암-관련된 돌연변이의 검출율을 나타내고 잠재적인 암-관련된 돌연변이를 나타내기 위한 기준으로서 4개 및 6개의 발생(r)을 사용하는 그래프(600)이다. 동일한 r 을 사용하여, 혈장 속에서 종양-기원한 DNA의 분획 농도가 높을 수록 혈장 속에서 검출가능한 암-관련된 돌연변이의 보다 많은 수가 생성될 수 있다. 혈장 속에서 종양-기원한 DNA의 동일한 분획 농도를 사용하여, r 이 높을 수록 검출된 돌연변이의 수는 더 작게 생성될 수 있다.

[0125] B. 오차로 인해 단일 검출된 거짓-양성의 수

[0126] 혈장 DNA 서열분석 데이터에 있어서 단일 뉴클레오타이드 변화는 서열분석 및 정렬 오차로 인하여 발생할 수 있다. 거짓-양성 단일 뉴클레오타이드 변화를 지닌 뉴클레오타이드 위치의 수는 이항 분포를 기초로 하여 수학적으로 예측할 수 있다. 거짓-양성 부위(N_{FP})의 수에 영향을 미치는 매개변수는 다음의 (1) 내지 (4)를 포함할 수 있다: (1) 서열분석 오차율(E) - 서열분석 오차율은, 서열분석된 뉴클레오타이드의 비율이 부정확한 것으로서 정의된다; (2) 서열분석 깊이(D) - 서열분석 깊이가 높을 수록, 서열분석 오차를 나타내는 뉴클레오타이드 위치의 수는 증가할 수 있다; (3) 잠재적인 암-관련된 돌연변이(r)을 정의하기 위한 동일한 뉴클레오타이드 변화의 최소의 발생 수; 및 (4) 목적인 영역 내에서 뉴클레오타이드 위치의 총 수(N_T).

[0127] 돌연변이의 발생은 일반적으로 무작위 공정으로서 고려될 수 있다. 따라서, 잠재적인 돌연변이를 정의하기 위한 발생 기준이 증가하면, 거짓 양성 뉴클레오타이드 위치의 수는 r 과 함께 기하급수적으로 감소할 수 있다. 존재하는 서열분석 플랫폼들 중 일부에서, 특정의 서열 내용은 서열분석 오차를 보다 더 갖는 경향이 있다. 이러한 서열분석 내용의 예는 GGC 모티프(motif), 단독중합체(예를 들면, AAAAAA), 및 단순한 반복체(예를 들면, ATATATATAT)를 포함한다. 이들 서열 내용은 단일의 뉴클레오타이드 변화 또는 삽입/결실 인공물을 실질적으로 증가시킬 것이다(참조: Nakamura K 등 Nucleic Acids Res 2011;39,e90 및 Minoche AE 등 Genome Biol 2011; 12,R112). 또한, 단독중합체 및 단순한 반복체와 같은 반복 서열은 정렬 속에 모호성을 컴퓨터적으로 도입시킬 수 있으므로, 단일의 뉴클레오타이드 변이에 대한 거짓-양성 결과를 초래한다.

[0128] 목적인 영역이 클수록, 관찰될 수 있는 거짓-양성 뉴클레오타이드 위치의 수가 더 커진다. 전체 게놈에서 돌연변이를 찾는 경우, 목적인 영역은 이의 게놈일 수 있으며, 포함된 뉴클레오타이드의 수는 3십억일 수 있다. 한편, 엑손에 초점을 맞추는 경우, 당해 엑손을 암호화하는 뉴클레오타이드의 수, 즉, 대략 4500만개가 목적인 영역을 구성할 수 있다.

[0129] 서열분석 오차와 관련된 거짓-양성 뉴클레오타이드 위치의 수는 다음의 계산을 기준으로 측정할 수 있다. 서열분석 오차로 인하여 동일한 위치에서 동일한 뉴클레오타이드 변화를 가질 확률(P_{Er})은 다음과 같이 계산할 수 있다:

$$P_{Er} = C(D, r)E \left(\frac{E}{3} \right)^{r-1}$$

[0130]

[0131] 상기 식에서, $C(D, r)$ 은 전체 D 성분으로부터 r 성분을 선택하기 위한 가능한 조합의 수이고; r 은 잠재적인 돌연변이를 정의하기 위한 발생의 수이며; D 는 서열분석 깊이이고; E 는 서열분석 오차율이다. $C(D, r)$ 은 다음과 같이 계산될 수 있다:

$$C(D, r) = \frac{D!}{r!(D-r)!}$$

[0132]

[0133] 돌연변이에 대해 거짓-양성인 뉴클레오타이드 위치의 수(N_{FP})는 다음과 같이 계산될 수 있다.

$$N_{FP} = N_I P_{Er}$$

[0134]

[0135] 상기 식에서, N_I 는 목적인 영역 내 뉴클레오타이드 위치의 전체 수이다.

[0136] 도 6b는 4, 5, 6 및 7 대 서열분석 깊이의 발생 기준(r)을 사용하여 뉴클레오타이드 변화를 갖는 것으로 거짓 분류된 뉴클레오타이드 위치의 예측된 수를 나타내는 그래프 (650)이다. 목적인 영역은 당해 계산에서 전체 게놈(30억개 뉴클레오타이드 위치)인 것으로 추정된다. 서열분석 오차율은 서열분석된 뉴클레오타이드의 0.3%인 것으로 추정된다. 알 수 있는 바와 같이, r 의 값은 거짓 양성에 유의적인 영향을 갖는다. 그러나, 도 6a로부터 알 수 있는 바와 같이, 검출된 돌연변이의 수를 감소시키는 r 의 값이 높을 수록, 적어도 유의적으로 될때까지 사용되는 서열분석 깊이는 더 높다.

[0137] C. 최소 발생(r)의 선택

[0138] 위에서 논의한 바와 같이, 서열분석 오차로 인한 실제 암 관련된 돌연변이 부위 및 거짓 양성 부위의 수는 서열분석 깊이와 함께 증가될 수 있다. 그러나, 이들의 증가율은 상이할 수 있다. 따라서, 서열분석 깊이 및 r 의 값의 선택을 사용하여 낮은 값에서 거짓 양성 부위의 수를 유지하면서 실제 암-관련된 돌연변이의 검출을 최대화하는 것이 가능하다.

[0139] 도 7a는 상이한 서열분석 깊이를 지닌 거짓 양성 부위 및 실제 암 관련된 돌연변이 부위의 수를 나타내는 그래프 (700)이다. 종양 조직에서 암 관련된 돌연변이의 총 수는 3,000인 것으로 추정되며 혈장 속에서 종양-기원한 DNA의 분획 농도는 10%인 것으로 추정된다. 서열분석 오차율은 0.3%인 것으로 추정된다. 궁극적으로, TP는, 상응하는 돌연변이가 종양 조직 속에 존재하는 실제-양성 부위를 나타내고, FP는 종양 조직 내 상응하는 돌연변이가 존재하지 않는 거짓-양성 부위를 나타내며 서열분석 데이터에 존재하는 뉴클레오타이드 변화는 서열분석 오차에 기인한다.

[0140] 그래프 (700)으로부터, 110배의 서열분석 깊이에서, 본 발명자가 기준으로서 6의 최소 발생($r=6$)을 사용하여 혈장 속의 잠재적인 돌연변이 부위를 정의하는 경우 1,410개의 실제 암 관련된 돌연변이가 검출될 수 있다. 당해 기준을 사용하면, 대략 20개의 거짓 양성 부위만이 검출될 수 있다. 본 발명자가 잠재적인 돌연변이를 정의하기 위한 기준으로서 최소 7의 발생($r=7$)을 사용하는 경우, 검출될 수 있는 암 관련된 돌연변이의 수는 대략 940으로 470까지 감소될 수 있다. 따라서, $r=6$ 의 기준은 혈장 내 암 관련된 돌연변이의 검출이 보다 민감성이라도 할 수 있다.

[0141] 한편, 200배의 서열분석 깊이에서, 검출된 실제 암 관련된 돌연변이의 수는, 본 발명자가 각각 6 및 7의 최소 발생(r)의 기준을 사용하여 잠재적인 돌연변이를 정의하는 경우, 대략 2,800 내지 2,600일 수 있다. 이들 r 의 2개 값을 사용하면, 거짓 양성 부위의 수는 각각 대략 740 및 20이 될 수 있다. 따라서, 200배의 서열분석 깊이에서, 잠재적인 돌연변이를 정의하기 위한 $r=7$ 의 보다 엄격한 기준의 사용은 실제 암-관련된 돌연변이를 검출하기 위한 민감성에 있어 유의적인 부작용없이 거짓-양성 부위의 수를 크게 감소시킬 수 있다.

[0142] D. 혈장 속에서 잠재적인 돌연변이를 정의하기 위한 서열분석 데이터에 대한 역학적 컷오프

[0143] 목적인 영역 내 각각의 뉴클레오타이드의 서열분석 깊이는 상이할 수 있다. 본 발명자들이 혈장 내 잠재적 돌연변이를 정의하기 위한 뉴클레오타이드 변화의 발생을 위한 고정된 컷오프 값을 적용하는 경우, 더 많은 서열 판독치에 의해 포괄되는 뉴클레오타이드(즉, 높은 서열분석 깊이)는 보다 낮은 서열분석 깊이를 가진 뉴클레오타이드와 비교한 서열분석 오차로 인하여 종양 조직 내 이러한 변화의 부재시 뉴클레오타이드 변이를 가진 것으로 거짓으로 표지될 확률이 더 높을 수 있다. 이러한 문제를 극복하기 위한 한가지 구현에는 특수 뉴클레오타이드 위치의 실제 서열분석 깊이에 따라서 및 거짓-양성 변이를 요청하기 위한 확률의 바람직한 상한치에 따라서 상이한 뉴클레오타이드 위치에 r 의 역학적 컷오프 값을 적용하는 것이다.

[0144] 하나의 구현예에서, 최대 허용가능한 거짓 양성률은 1.5×10^8 개 뉴클레오타이드 위치중 1에서 고정될 수 있다. 이러한 최대 허용가능한 거짓-양성률을 사용하여, 전체 게놈 내에서 확인되는 거짓-양성 부위의 총 수는 20 미만일 수 있다. 상이한 서열분석 깊이에 대한 r 의 값은 도 6b에 나타난 그래프에 따라 측정할 수 있으며 이들 컷오프는 표 1에 나타낸다. 다른 구현예에서, 다른 상이한 최대 허용가능한 거짓 양성률, 예를 들면, 3×10^8

개중 1개 또는 6×10^7 개 중 1개를 사용할 수 있다. 거짓-양성 부위의 상응하는 총 수는 각각 10, 30 및 50개 미만일 수 있다.

표 1

[0145] 특수한 뉴클레오타이드 위치의 상이한 서열분석 깊이에 대한 잠재적인 돌연변이(r)를 정의하기 위해 혈장 속에 존재하는 뉴클레오타이드 변화의 발생의 최소 수. 최대 거짓 양성률은 1.5×10^8 개 뉴클레오타이드중 1개에서 고정된다.

특수한 뉴클레오타이드 위치의 서열분석 깊이	잠재적 돌연변이(r)를 정의하기 위한 혈장 DNA 서열분석 데이터 속에 존재할 뉴클레오타이드 변화의 발생의 최소 수
<50	5
50-110	6
111-200	7
201-310	8
311-450	9
451-620	10
621-800	11

E. 표적 농축 서열분석

[0147] 도 7a에 나타난 바와 같이, 보다 높은 서열분석 깊이는 r의 보다 높은 값의 사용을 허용함으로써 다수의 거짓 양성 부위를 낮게 유지하면서 암 관련된 돌연변이를 검출하기 위한 우수한 민감성을 생성할 수 있다. 예를 들면, 110배의 서열분석 깊이에서, 1,410개의 실제 암 관련된 돌연변이가 6의 r 값을 사용하여 혈장 속에서 검출될 수 있는 반면, 검출된 실제 암 관련된 돌연변이의 수는, 서열 분석 깊이가 200배 증가하고 7의 r 값이 적용되는 경우 2,600일 수 있다. 2세트의 데이터는 대략 20의 거짓 양성 부위의 예측된 수를 제공할 수 있다.

[0148] 200배의 깊이에 대한 전체 게놈의 서열분석은 현재 비교적 비싸지만, 이러한 서열분석 깊이를 달성하는 한가지 가능한 방법은 목적인 보다 작은 영역에 초점을 맞추는 것일 수 있다. 표적 영역의 분석은 예를 들면, 하이브리드화에 의해 목적인 게놈 영역을 포획하기 위하여 DNA 또는 RNA 미끼를 사용함으로써 달성할 수 있지만, 이에 한정되지 않는다. 이후에, 포획된 영역은 예를 들면, 자기적 수단(magnetic means)에 의해 제거하여 서열분석에 적용시킨다. 이러한 표적 포획은 예를 들면, Agilent SureSelect 표적 농축 시스템, Roche Nimblegen 표적 농축 시스템 및 Illumina 표적화된 재서열분석 시스템을 사용하여 수행할 수 있다. 다른 시도는 표적 영역의 PCR 증폭을 수행한 후 서열분석을 수행하는 것이다. 하나의 구현예에서, 목적인 영역은 진유전체(exome)이다. 이러한 구현예에서, 모든 엑손의 표적 포획은 혈장 DNA에서 수행하며, 엑손 영역에 대해 농축된 혈장 DNA를 이후 서열분석할 수 있다.

[0149] 보다 높은 서열분석 깊이를 갖는 것 외에도, 전체 게놈을 분석하는 대신 구체적인 영역에 초점을 맞추는 것은 조사 공간에서 뉴클레오타이드 위치의 수를 유의적으로 감소시킬 수 있으며 동일한 서열분석 오차율을 제공한 거짓 양성 부위의 수에 있어서의 감소를 초래할 수 있다.

[0150] 도 7b는 전체 게놈(WG) 및 모든 엑손의 분석을 포함하는 거짓 양성 부위의 예측된 수를 나타내는 그래프 (750)이다. 분석의 각각의 유형의 경우, r에 대해 2개의 상이한 값, 5 및 6이 사용된다. 200배의 서열분석 깊이에서 r=5가 사용되어 혈장 속의 돌연변이를 정의하는 경우, 거짓 양성 부위의 예측된 수는 전체 게놈 및 모든 엑손에 대해 각각 대략 23,000 및 230이다. r=6이 사용되어 혈장 속의 돌연변이를 정의하는 경우, 거짓 양성 부위의 예측된 수는 각각 750 및 7이다. 따라서, 목적인 영역 내 뉴클레오타이드의 수의 한계는 혈장 돌연변이 분석에서 거짓 양성률의 수를 유의적으로 감소시킬 수 있다.

[0151] 엑손 포획 또는 심지어 진유전체 포획 서열분석시, 조사 공간에서 뉴클레오타이드의 수는 감소된다. 따라서, 본 발명자가 암-관련된 돌연변이의 검출을 위한 보다 높은 거짓-양성율을 허용하는 경우, 거짓-양성 부위의 절대값은 비교적 낮은 수준으로 유지될 수 있다. 보다 높은 거짓-양성율의 허용은 사용될 혈장 속의 단일 뉴클레오타이드 변이를 정의하기 위해 거의 엄격하지 않은 최소 발생 기준(r)을 허용할 수 있다. 이는 실제 암-관련된 돌연변이의 검출을 위해 보다 높은 민감성을 생성할 수 있다.

[0152] 하나의 구현예에서, 본 발명자는 1.5×10^6 의 최대 허용가능한 거짓 양성률을 사용할 수 있다. 당해 거짓 양성

를 사용하면, 표적화된 엑손내 거짓 양성 부위의 전체 수는 단지 20일 수 있다. 1.5×10^6 의 최대 허용가능한 거짓 양성률을 사용하여 상이한 서열분석 깊이에 대한 r의 값은 표 2에 나타낸다. 다른 구현예에서, 다른 상이한 최대 허용되는 거짓 양성률, 예를 들면, 3×10^6 중 1, 10^6 중 1 또는 6×10^5 중 1이 사용될 수 있다. 거짓-양성 부위의 상응하는 전체 수는 각각 10, 30 및 50 미만일 수 있다. 하나의 구현예에서, 상이한 부류의 돌연변이는 위에서 기술한 바와 같이 상이한 청량에 기여할 수 있다.

표 2

[0153] 특수한 뉴클레오타이드 위치의 상이한 서열 서열분석 깊이에 대한 잠재적인 돌연변이(r)를 정의하기 위해 혈장 속에 존재하는 뉴클레오타이드 변화의 최소 발생 수. 최대 거짓-양성율은 1.5×10^6 개 뉴클레오타이드중 1로 고정된다.

특수한 뉴클레오타이드 위치의 서열 분석 깊이	잠재적인 돌연변이를 정의하기 위한 혈장 DNA 서열분석 데이터에 존재할 뉴클레오타이드 변화의 최소 발생 수(r)
<50	4
50-125	5
126-235	6
236-380	7
381-560	8
561-760	9

[0154] VIII. 암 검출

[0155] 위에서 논의한 바와 같이, 변이체 유전자좌에서 서열 태그의 수를 다양한 방법으로 사용하여 매개변수를 결정할 수 있고, 이를 암의 수준을 분류하기 위해 역치와 비교한다. 유전자좌 또는 많은 유전자좌에서 모든 관독물에 대한 변이체 관독물의 분획 농도는 사용될 수 있는 다른 매개변수이다. 하기는 매개변수 및 역치를 계산하는 일부 예이다.

[0156] A. 매개변수의 측정

[0157] CG가 첫 번째 대립형질에 대한 특수 유전자좌에서 동형접합성이고 변이체 대립형질이 생물학적 시료(예를 들면, 혈장) 속에서 관찰되는 경우, 분획 농도는 $2p / (p+q)$ 로 계산될 수 있으며, 여기서 p는 변이체 대립형질을 갖는 서열 태그의 수이고 q는 CG의 첫 번째 대립형질을 갖는 서열 태그의 수이다. 당해 식은, 종양의 반수체형중 단지 하나가 변이체를 갖는 것으로 추정하는데, 이것이 전형적인 경우일 수 있다. 따라서, 각각의 동형접합성 유전자좌에 대해 분획 농도가 계산될 수 있다. 분획 농도는 평균낼 수 있다. 다른 구현예에서, 총수 p는 모든 유전자좌, 및 유사하게 총수 q의 경우 다수의 서열 태그를 포함함으로써 분획 농도를 측정할 수 있다. 이제 예를 기술한다.

[0158] 4명의 HCC 환자의 혈장 속에서 종양 기원한 단일의 뉴클레오타이드 변이체(SNV)의 게놈 전체 검출을 실험하였다. 본 발명자는 종양 DNA 및 연막(buffy coat) DNA를 각각 29.5배(범위, 27배 내지 33배) 및 43배(범위, 39배 내지 46배)의 반수체 게놈 포함의 평균 깊이로 서열분석하였다. 4명의 HCC 환자 각각으로부터 종양 DNA 및 연막 DNA로부터의 MPS 데이터를 비교하고, 종양 DNA 속에 존재하지만 연막 DNA 속에는 존재하지 않는 SNV를 엄격한 생물정보학자 알고리즘으로 조사하였다. 당해 알고리즘은, 이것이 실제 SNV로서 분류될 수 있기 전에 서열분석된 종양 DNA 단편의 적어도 역치 수(즉, 상응하는 서열분석된 태그내) 속에 존재할 추정의 SNV를 필요로 하였다. 당해 역치수는 특수한 뉴클레오타이드의 서열분석 깊이 및 예를 들면, 본원에 기술된 것으로서, 서열분석 오차율을 고려하여 결정하였다.

[0159] 도 8은 본 발명의 구현예에 따라 혈장 속에서 종양 유래 DNA의 분획 농도를 포함하는, 치료 전 및 후에 4명의 HCC 환자에 대한 결과를 나타내는 표 (800)이다. 종양-기원한 SNV의 수는 4명의 HCC 경우에 1,334 내지 3,171의 범위였다. 혈장 속에서 검출가능했던 이러한 SNV의 비율은 치료 전 후에 나열한다. 치료 전에, 종양 관련된 SNV의 15% 내지 94%가 혈장 속에서 검출되었다. 치료 후, 퍼센트는 1.5% 내지 5.5%이었다. 따라서, 검출된 SNV의 수는 암의 수준과 상호 관련된다. 이는, SNV의 수를 매개변수로 사용하여 암의 수준을 분류할 수 있음을 나타낸다.

[0160] 혈장 속에서 종양 유래 DNA의 분획 농도는 전체(즉, 돌연변이체 및 야생형) 서열과 관련하여 돌연변이체의 분획

계수로 결정하였다. 식은 $2p / (p+q)$ 이며, 여기서 2는 종양에서 돌연변이되는 단지 하나의 반수체에 대해 계수된다. 이들 분획 농도는 게놈전체의 응집된 대립형질 손실(GAAL) 분석으로 측정된 것과 잘 관련되었으며(참조: Chan KC 등 Clin Chem 2013;59:21 1-24) 수술 후 감소되었다. 따라서, 분획 농도는 또한 암의 수준을 결정하기 위한 이용가능한 매개변수인 것으로 밝혀져 있다.

[0161] SNV 분석으로부터의 분획 농도는 종양 부하를 전달할 수 있다. 보다 높은 종양 부하(예를 들면, 보다 높은 유추된 분획 농도)를 갖는 암 환자는 보다 낮은 종양 부하를 갖는 환자보다 보다 높은 빈도의 체세포 돌연변이를 가질 것이다. 따라서, 구현예를 또한 예후에 사용할 수 있다. 일반적으로, 보다 높은 종양 부하를 갖는 암 환자들은 보다 낮은 종양 부하를 가진 환자들보다 불량한 예후를 갖는다. 암 환자 그룹은 따라서 당해 질병으로부터 사망할 기회가 더 높다. 일부 구현예에서, 생물학적 시료, 예를 들면, 혈장 속에서 DNA의 절대 농도를 측정 후(예를 들면, 실시간 PCR 또는 형광분석법), 종양-관련된 유전적 이상의 절대 농도를 결정하여 임상 검출 및/또는 모니터링 및/또는 예후에 사용할 수 있다.

[0162] B. 역치의 측정

[0163] 표 800은 역치를 결정하기 위해 사용될 수 있다. 앞서 언급한 바와 같이, SNV의 수 및 SNV 분석에 의해 측정된 분획 농도는 암의 수준과 관련된다. 당해 역치는 개인을 기준으로 측정할 수 있다. 예를 들어, 전 처리값을 사용하여 역치를 측정할 수 있다. 각종 시행에서, 역치는 절대 값의 전 처리로부터의 상대적인 변화일 수 있다. 적합한 역치는 SNV의 수 또는 분획 농도에 있어 50%까지의 감소일 수 있다. 이러한 역치는 표 800에서 경우들 각각에 대해 보다 낮은 수준의 암의 분류를 제공할 수 있다. 이러한 역치는 서열분석 깊이에 의존적일 수 있음에 주목한다.

[0164] 하나의 구현예에서, 역치는 시료에 거쳐 사용할 수 있으며, 매개변수에 대한 전-처리값을 계수하거나 계수하지 않는다. 예를 들어, 100개의 SNV의 역치를 사용하여 피검자를 암이 없거나 낮은 수준의 암을 가진 것으로 분류할 수 있다. 100개의 SNV의 이러한 역치는 표 800에서 4개의 경우 각각에 의해 만족된다. 분획 농도가 매개변수로서 사용된 경우, 1.0%의 역치는 HCC1 내지 HCC3를 실제로 0 수준의 암으로서 분류할 수 있으며, 1.5%의 제2 역치는 HCC4를 낮은 수준의 암으로 분류할 수 있다. 따라서, 1 이상의 역치를 사용하여 2 이상의 분류를 수득할 수 있다.

[0165] 다른 가능한 역치를 나열하기 위하여, 본 발명자는 종양-관련된 SNV에 대한 건강한 대조군의 혈장을 분석하였다. 다수의 측정을 건강한 피검자에게서 수행하여 구성적 게놈과 비교하여 생물학적 시료로부터 변이가 예측되는 범위를 측정할 수 있다.

[0166] 도 9는 본 발명의 구현예에 따라서 16명의 건강한 대조군 피검자에서 HCC 관련된 SNV의 검출을 나타내는 표 900이다. 표 900을 사용하여 SNV 분석 시도의 특이성을 추정할 수 있다. 16명의 건강한 대조군은 상이한 열로 나열된다. 컬럼은 특이적인 HCC 환자에 대해 검출된 SNV를 시험하며 변이체 대립형질을 갖는 변이체 유전자좌에서 서열 관독물의 수 및 야생형 대립형질(즉, CG로부터의 대립형질)을 지닌 서열 관독물의 수를 나타낸다. 예를 들면, HCC1의 경우, 대조군 C01은 이러한 변이체 유전자좌에서 40개의 변이체 관독물을 가졌지만, 야생형 대립형질중 31,261개의 관독물을 가졌다. 마지막 컬럼은 HCC1 환자에 대한 SNV 모두에 따라 전체 분획 농도를 나타낸다. HCC 관련된 SNV는 HCC 환자에 대해 특이적이었으므로, HCC 관련된 SNV의 존재는 거짓 양성을 나타낸다. 본원에 기술된 것으로서, 컷오프 값을 이들 명확한 서열 변이체에 적용하는 경우, 이들 거짓 양성 모두는 필터링될 수 있다.

[0167] 16명의 건강한 대조군의 혈장 속에서 이들 추정된 종양 관련된 돌연변이의 작은 수의 존재는 당해 방법의 통계학적 노이즈(stochastic noise)"를 나타내었으며 서열 오차에 기인하는 경향이 있었다. 이러한 노이즈로부터 평가된 평균 분획 농도는 0.38%이었다. 이들 값은 건강한 피검자에 대한 범위를 나타낸다. 따라서, HCC에 대한 0 수준의 암의 분류를 위한 역치 값은, 최대 분획 농도가 0.43%이었으므로, 약 0.5%일 수 있다. 따라서, 모든 암 세포가 HCC 환자로부터 제거되는 경우, 이들 낮은 분획 농도가 예측될 수 있다.

[0168] 표 800을 다시 참조하면, 0.5%를 0 수준의 암에 대한 역치로서 사용한 경우, HCC1 및 HCC3에 대한 처리 후 혈장 데이터는 SNV 분석을 기준으로 0 수준을 갖는 것으로 측정될 수 있다. HCC2는 0으로부터 1 수준 상향으로 분류될 수 있다. HCC4는 또한 0으로부터 1 수준 상향, 또는 일부 더 높은 수준으로 분류될 수 있지만 처리 전 시료와 비교하여 여전히 상대적으로 낮은 수준일 수 있다.

[0169] 매개변수가 변이체 유전자좌의 수에 상응하는 하나의 구현예에서, 역치는 0일 수 있다(즉, 하나의 변이체 유전자좌는 0이 아닌 수준의 암을 나타낼 수 있다). 그러나, 많은 셋팅(예를 들면, 깊이의)에서, 역치는 더 높을

수 있는데, 예를 들면, 5 또는 10의 절대값일 수 있다. 개인을 처리 후 모니터링하는 하나의 시행에서, 역치는 시료 속에서 나타나는 SNV의 특정 퍼센트(종양을 직접 분석함으로써 확인됨)일 수 있다. 유전자좌에서 요구된 변이체 판독물의 수에 대한 컷오프 값이 충분히 큰 경우, 하나의 변이체 유전자좌를 갖는 것으로도 0이 아닌 수준의 암의 지표일 수 있다.

[0170] 따라서, 생물학적 시료(예를 들면, 혈장)로부터의 DNA중 변이(예를 들면, 단일 뉴클레오타이드 변이)의 정량적 분석을 암의 진단, 모니터링 및 예후에 사용할 수 있다. 암의 검출을 위해, 시험한 피검자의 혈장 속에서 검출된 다수의 단일의 뉴클레오타이드 변이를 건강한 피검자의 그룹의 것과 비교할 수 있다. 건강한 피검자에서, 혈장 속의 명백한 단일의 뉴클레오타이드 변이는 혈액 세포 및 다른 기관으로부터의 서열분석 오차, 비-클론성 돌연변이로 기인할 수 있다. 정상인 건강한 피검자에서 세포는 표 900에 나타낸 바와 같이, 소수의 돌연변이를 수반할 수 있는 것으로 밝혀졌다(참조: Conrad DF 등 Nat Genet 2011;43:712-4). 따라서, 명확하게 건강한 피검자의 그룹의 혈장 속의 명백한 단일 뉴클레오타이드 변이의 전체 수는, 시험한 환자가 0이 아닌 수준의 암에 상응하는 혈장 속의 비정상적으로 높은 수의 단일 뉴클레오타이드 변이를 가지는지를 결정하기 위한 참조 범위로 사용될 수 있다.

[0171] 참조 범위를 확립하는데 사용된 건강한 피검자는 연령 및 성별의 측면에서 시험한 피검자와 조화될 수 있다. 앞서의 연구에서, 체세포내 다수의 돌연변이는 연령에 따라 증가할 수 있는 것으로 밝혀졌다(참조: Cheung NK 등, JAMA 2012;307:1062-71). 따라서, 나이가 들어감에 따라, 이들이 비교적 대부분의 시점에 양성이거나 임상적으로 유의적인 것이 될 때까지 매우 긴 시간이 걸릴 수 있다고 해도, 세포의 클론을 추적하는 것이 '정상'이 될 수 있다. 하나의 구현예에서, 참조 수준은 상이한 피검자 그룹, 예를 들면, 상이한 연령, 성별, 민족성, 및 다른 매개변수(예를 들면, 흡연 상태, 간 상태, 알코올, 약물 복용력)에 대해 생성될 수 있다.

[0172] 참조 범위는 사용된 컷오프 값(즉, 유전자좌에서 요구된 변이체 서열 태그의 수), 및 또한 추정된 거짓 양성률 및 다른 변수(예를 들면, 연령)를 기초로 하여 변할 수 있다. 따라서, 참조 범위는 하나 이상의 기준의 특수 세트에 대해 측정될 수 있으며, 동일한 기준을 사용하여 시료에 대한 매개변수를 결정할 수 있다. 이후에, 매개변수를 참조 범위와 비교할 수 있는데, 이는 이들 둘 다를 동일한 기준을 사용하여 결정하였기 때문이다.

[0173] 위에서 언급한 바와 같이, 구현예는 암의 수준을 결정하기 위한 다수의 역치를 사용할 수 있다. 예를 들어, 제1 수준은 역치 미만의 매개변수, 및 예비-신생물 수준일 수 있는, 암의 적어도 제1 수준에 대한 암의 신호를 측정할 수 없다. 다른 수준은 암의 상이한 단계에 상응할 수 있다.

[0174] C. 실험적 변수에 대한 의존성

[0175] 서열분석 깊이는 소수(예를 들면, 종양) 게놈의 최소 검출 역치를 확립하는데 중요할 수 있다. 예를 들어, 10개 반수체 게놈의 서열분석 깊이를 사용하는 경우, 어떠한 오차없이 서열분석 기술을 사용하는 경우에도 검출할 수 있는 최소의 종양 DNA 농도는 1/5, 즉 20%이다. 다른 한편, 100개의 반수체 게놈의 서열분석 깊이를 사용하는 경우, 2%로 감소시킬 수 있다. 당해 분석은, 단지 하나의 돌연변이 유전자좌가 분석되는 시나리오로 언급된다. 그러나, 보다 많은 돌연변이 유전자좌가 분석되는 경우, 최소의 종양 DNA 농도는 더 낮아질 수 있으며 이항 확률 함수에 의해 통제된다. 예를 들어, 서열분석 깊이가 10배이고 종양 DNA의 분획 농도가 20%인 경우, 돌연변이를 검출하는 기회는 10%이다. 그러나, 10개의 돌연변이를 갖는 경우, 적어도 하나의 돌연변이를 검출하는 기회는 $1 - (1 - 10\%)^{10} = 65\%$ 일 수 있다.

[0176] 서열분석 깊이를 결정하기 위한 수개의 실행이 존재한다. 서열분석 깊이가 클 수록, 보다 많은 서열분석 오차가 관찰될 수 있다(참조: 도 4 및 5). 그러나, 서열분석 깊이가 클 수록, 서열분석 오차가 게놈에서 무작위적으로 발생하지만 돌연변이는 제공된 세포 집단에 대해 동일한 위치에서 발생할 수 있으므로 세포(예를 들면, 암 세포)의 소집단의 클론성 확장으로 인하여 돌연변이로부터 서열분석 오차를 보다 용이하게 차별화시킬 수 있다.

[0177] 서열분석 깊이가 클 수록, "건강한 세포"로부터 보다 많은 돌연변이가 확인될 수 있다. 그러나, 이러한 건강한 세포의 클론성 확장이 존재하지 않고 이들의 돌연변이성 프로파일이 상이한 경우, 이들 건강한 세포 속에서 돌연변이는 혈장 속에서 이들의 발생 빈도로 인하여 돌연변이로부터 차별화될 수 있다(예를 들면, N = 2, 3, 4, 5 이상과 같은, 돌연변이를 나타내는 요구되는 수의 판독물에 대한 컷오프 N을 사용함에 의한).

[0178] 위에서 언급한 바와 같이, 역치는 클론적으로 확장될 수 있는 건강한 세포 내 돌연변이의 양에 의존할 수 있으므로, 다른 메커니즘을 통해 필터링될 수 없다. 예측할 수 있는 이러한 변이는 건강한 피검자를 분석함으로써 수득할 수 있다. 클론성 확장이 시간에 따라 발생하므로, 환자의 연령은 건강한 피검자에서 관찰되는 변이에

영향을 미칠 수 있으므로, 역치는 연령에 의존할 수 있다.

[0179]

D. 표적화된 시도와의 조합

[0180]

일부 구현예에서, 무작위 서열분석을 표적화된 시도와 함께 사용할 수 있다. 예를 들면, 암 환자의 표시 시 혈장 시료의 무작위 서열분석을 수행할 수 있다. 혈장 DNA의 서열분석 데이터는 카피 수 일탈 및 SNV에 대해 분석될 수 있다. 일탈(예를 들면, SNV의 증폭/결실 또는 고 밀도)을 나타내는 영역을 일련의 모니터링 목적을 위해 표적화할 수 있다. 모니터링은 시간에 따라 수행하거나 단일 과정으로서 효과적으로, 무작위 서열분석 직후 수행할 수 있다. 표적화된 분석을 위해, 용액 상 하이브리드화계 포획 시도를 성공적으로 사용하여 비침입성 태아 진단을 위해 혈장 DNA를 농축시켜 왔다(참조: Liao GJ 등 Clin Chem 2011 ;57:92-101). 이러한 기술은 위에 언급되어 있다. 따라서, 표적화되고 무작위적인 시도는 암 검출 및 모니터링을 위한 조합에서 사용할 수 있다.

[0181]

따라서, 표적화되지 않은, 상기 언급한 게놈전체 시도를 사용하여 잠재적으로 돌연변이되는 것으로 밝혀진 유전자좌의 표적화된 서열분석을 수행할 수 있다. 이러한 표적화된 서열분석은 용액- 또는 고체-상 하이브리드화 기술(예를 들면, Agilent SureSelect, NimbleGen Sequence Capture, 또는 Illumina 표적화된 재서열분석 시스템을 사용)을 사용한 후 거대한 평행 서열분석으로 수행할 수 있다. 다른 시도는 표적화된 서열분석을 위한 증폭(예를 들면, PCR 계) 시스템을 수행하는 것이다(참조: Forshew T 등 Sci Transl Med 2012; 4: 135ra68).

[0182]

IX. 분획 농도

[0183]

종양 DNA의 분획 농도를 사용하여 유전자좌를 돌연변이로서 확인하기 전에 유전자좌에서 변이의 요구되는 수에 대한 컷오프 값을 측정할 수 있다. 예를 들어, 분획 농도가 비교적 높은 것으로 밝혀진 경우, 높은 컷오프를 사용하여 보다 많은 거짓 양성을 필터링할 수 있는데, 이는, 비교적 높은 수의 변이체 판독물이 실제 SNV에 대해 존재할 수 있음을 알고 있기 때문이다. 한편, 분획 농도가 낮은 경우, 보다 낮은 컷오프가 요구됨으로써 일부 SNV는 소실되지 않는다. 이 경우, 분획 농도는 SNV 분석보다는 상이한 방법에 의해 측정할 수 있으며, 여기서 이는 매개변수로서 사용된다.

[0184]

각종 기술을 분획 농도를 결정하기 위해 사용할 수 있으며, 이들 중 일부가 본원에 기술되어 있다. 이들 기술은 혼합물, 예를 들면, 종양 세포 및 비악성 세포의 혼합물을 함유하는 생검 시료 또는 종양 세포로부터 방출된 DNA 및 비악성 세포로부터 방출된 DNA를 함유하는 암 환자로부터의 혈장시료 중 종양 유래 DNA의 분획 농도를 결정하는데 사용할 수 있다.

[0185]

A. GAAL

[0186]

게놈전체 응집된 대립형질 손실(genomewide aggregated allelic loss: GAAL)은 이형접합성을 손실한 유전자좌를 분석한다(참조: Chank C 등 Clin Chem 2013;59:21 1-24). 이형접합성인 구성적 게놈 CG의 부위의 경우, 종양은 종종 대립형질 중 하나의 결실을 갖는 유전자좌를 갖는다. 따라서, 이러한 유전자좌에 대한 서열 판독물은 다른 것 보다 하나 이상의 대립형질을 나타낼 것이며, 여기서 이러한 차이는 시료 속의 종양 DNA의 분획 농도에 비례한다. 이러한 계산의 예는 다음과 같다.

[0187]

HCC 환자의 연막 및 종양 조직으로부터 추출된 DNA는 Affymetrix Genome- Wide Human SNP Array 6.0 system을 사용하여 유전자형을 분석하였다. 미세배열 데이터를 Affymetrix Genotyping Console version 4.1로 프로세싱하였다. 유전형 분석 및 단일-뉴클레오타이드 다형체(SNP) 호출은 Birdseed v2 알고리즘을 사용하여 수행하였다. 연막 및 종양 조직에 대한 유전형 데이터는 이형접합성의 손실(loss-of-heterozygosity: LOH) 영역을 확인하고 카피 수 분석을 수행하는데 사용하였다. 카피 수 분석은 데플트 매개변수(공급원: Affymetrix) 및 100bp의 최소 게놈-분절 크기 및 분절내 최소의 5개 유전 마커를 사용한 Genotyping Console을 사용하여 수행하였다.

[0188]

LOH를 지닌 영역은 종양 조직 내에서 1개 카피 및 연막에서 2개 카피를 가지는 영역으로 확인되었으며, 이들 영역 내 SNP는 연막내에서 이형접합성이나 종양 조직 내에서는 동형접합성이었다. 종양 조직 내에서 LOH를 나타내는 게놈 영역의 경우, 연막에 존재하지만 종양 조직 내에서 부재하거나 강도가 감소된 SNP 대립형질은 염색체 영역의 결실된 분절에서 대립형질인 것으로 고려하였다. 연막 및 종양 조직 둘 다에 존재하는 대립형질은 염색체 영역의 결실되지 않은 분절로부터 기원하는 것으로 고려하였다. 종양 내 단일 카피 손실을 지닌 염색체 영역 모두의 경우, 결실된 대립형질 및 결실되지 않은 대립형질을 수반하는 서열 판독물의 전체 수를 계수하였다. 이들 2개 값의 차이를 사용하여 시료 속의 종양-기원한 DNA의 분획 농도(F_{GAAL})를 다음 식을 사용하여

추론하였다:

$$F_{GAAL} = \frac{N_{non-del} - N_{del}}{N_{non-del}}$$

[0189]

[0190]

상기 식에서, $N_{non-del}$ 는 결실되지 않은 대립형질을 지니는 서열 판독물의 총 수를 나타내고 N_{del} 은 결실된 대립형질을 지니는 서열 판독물의 총 수를 나타낸다.

[0191]

B. 게놈 표시를 사용한 평가

[0192]

GAAL 기술을 사용하는 문제점은, 특수한 유전자좌(즉, LOH를 나타내는 유전자좌)가 확인되고 이러한 유전자좌를 정렬하는 서열 판독물만이 사용된다는 것이다. 이러한 요건은 추가의 단계를 추가할 수 있으므로 비용소모적이다. 이제 카피 수, 예를 들면, 서열 판독물 밀도만을 사용하는 구현예를 기술한다.

[0193]

염색체 비정상, 예를 들면, 증폭 및 결실은 암 게놈 속에서 흔히 관찰된다. 암 조직에서 관찰된 염색체 비정상은 전형적으로 아염색체 영역을 포함하며 이러한 비정상은 1Mb보다 더 짧을 수 있다. 그리고, 암-관련된 염색체 비정상은 상이한 환자에서 이형접합성이므로 상이한 영역이 상이한 환자에서 영향받을 수 있다. 이는 또한 암 게놈에서 발견될 수식, 수백 또는 심지어 수천의 비정상의 경우 일반적이지 않다. 이들 인자 모두는 종양 DNA 농도를 결정하기 어렵게 만든다.

[0194]

양태는 종양-관련된 염색체 비정상으로부터 생성된 정량적 변화의 분석을 포함한다. 하나의 구현예에서, 암 세포 및 정상 세포로부터 기원한 DNA를 함유하는 DNA 시료를 거대한 평행 서열분석을 사용하여, 예를 들면, Illumina HiSeq2000 서열분석 플랫폼으로 서열분석한다. 유도된 DNA는 혈장 또는 다른 적합한 생물학적 시료 속에서 세포-유리된 DNA일 수 있다.

[0195]

종양 조직 속에서 증폭되는 염색체 영역은 서열분석되는 확률이 증가될 수 있고 종양 조직 속에서 결실되는 영역은 서열분석되는 확률이 감소될 수 있다. 그 결과, 증폭된 영역에 대해 정렬되는 서열 판독물의 밀도가 증가될 수 있고 결실된 영역에 대한 정렬은 감소될 수 있다. 변이 정도는 DNA 혼합물 속의 종양-기원한 DNA의 분획 농도에 비례한다. 종양 조직으로부터의 DNA의 비율이 높을 수록, 염색체 비정상에 의해 유발될 수 있는 변화가 더 커진다.

[0196]

1. 고 종양 농도를 지닌 시료 속에서 평가

[0197]

DNA를 4명의 간세포 암종 환자의 종양 조직으로부터 추출하였다. DNA를 Covaria DNA 초음파 시스템을 사용하여 분획화하고 기술한 바와 같이 Illumina HiSeq2000 플랫폼을 사용하여 서열분석하였다(참조: Chan C 등 Clin Chem 2013;59:211-24). 당해 서열 판독물을 사람 기준 게놈(hg18)에 대해 정렬하였다. 이후에, 게놈을 1 Mb 빈(영역)으로 분할하고 서열 판독물 밀도를 기술한 바와 같이 GC-편향에 대한 조절 후 각각의 빈에 대해 계산하였다(참조: Chen EZ 등 PLoS One. 2011;6:e21791).

[0198]

서열분석 판독물을 기준 게놈에 대해 정렬한 후, 서열 판독물 밀도를 다양한 영역에 대해 컴퓨터처리할 수 있다. 하나의 구현예에서, 서열 판독물 밀도는 기준 게놈(예를 들면, 기준 게놈 내 독특한 위치)에 대해 정렬될 수 있는 전체 서열 판독물로 나눈 특수한 빈(예를 들면, 1Mb 영역)에 대해 맵핑된 판독물의 수로서 측정된 비율이다. 종양 조직 속에서 증폭된 염색체 영역과 오우버랩된 빈은 이러한 중첩이 없는 빈으로부터의 것들 보다 더 높은 서열 판독물 밀도를 갖는 것으로 예측된다. 한편, 결실되는 염색체 영역과 중첩되는 빈은 이러한 중첩이 없는 것보다 더 낮은 서열 판독물 밀도를 갖는 것으로 예측된다. 염색체 비정상이 있는 영역과 없는 영역 사이의 서열 판독물 밀도에 있어서 차이의 크기는 시료 속에서 종양-기원한 DNA의 비 및 종양 세포 내 증폭/결실의 정도에 의해 주로 영향받는다.

[0199]

다양한 통계적 모델을 사용하여 상이한 유형의 염색체 비정상에 상응하는 서열 판독물 밀도를 갖는 빈을 확인할 수 있다. 하나의 구현예에서, 정상의 혼합물 모델(참조: McLachlan G and Peel D. Multivariate normal mixtures. In Finite mixture models 2004: p81-116. John Wiley & Sons Press)을 사용할 수 있다. 다른 통계적 모델, 예를 들면, 이항급수 혼합물 모델 및 푸아송 회귀 모델(참조: McLachlan G and Peel D. Mixtures with non-normal components, Finite mixture models 2004: p135- 174. John Wiley & Sons Press)을 또한 사용할 수 있다.

[0200]

빈에 대한 서열 판독물 밀도는 연막 DNA의 서열분석으로부터 측정된 바와 동일한 빈의 서열 판독물 밀도를 사용

하여 정규화할 수 있다. 상이한 빈의 서열 관독물 밀도는 특수한 염색체 영역의 서열 내용에 의해 영향받을 수 있으므로, 정규화는 비정상성을 나타내는 영역을 보다 정밀하게 확인하는데 도움을 줄 수 있다. 예를 들어, 상이한 염색체 영역의 맵확률(이는 서열을 이의 원래의 위치로 역 정렬하는 확률을 말한다)은 상이할 수 있다. 또한, 카피 수(즉, 카피 수 변이)의 다형체는 또한 빈의 서열 관독물 밀도에 영향을 미칠 수 있다. 따라서, 연막 DNA를 사용한 정규화는 상이한 염색체 영역 사이의 서열 내용에 있어서 차이와 관련된 변이를 잠재적으로 최소화시킬 수 있다.

[0201] 도 10a는 본 발명의 구현예에 따른 HCC 환자의 종양 시료의 서열 관독물 밀도의 분포 플롯 (1000)을 나타낸다. 종양 조직은 HCC 환자로부터의 외과적 절개 후 수득되었다. x-축은 종양 조직과 환자의 연막 사이의 서열 관독물 밀도의 비(R)의 \log_2 를 나타낸다. y-축은 빈의 수를 나타낸다.

[0202] 피크는 분포 곡선에 적합하도록 하여 정상의 혼합물 모델을 사용한 결실, 증폭을 지니고 염색체 비정상성이 없는 영역을 나타낼 수 있다. 하나의 구현예에서, 다수의 피크를 상이한 타당한 값에 걸친 아카이케 정보 기준 (Akaike's information criterion: AIC)으로 측정할 수 있다. $\log_2 R = 0$ (즉, $R = 1$)을 갖는 중심 피크는 어떠한 염색체 이상이 없는 영역을 나타낸다. 좌측 피크(중심 피크에 대해)는 하나의 카피가 손실된 영역을 나타낸다. 우측 피크(중심 피크에 대해)는 하나의 카피가 증폭된 영역을 나타낸다.

[0203] 종양 유래 DNA의 분획 농도는 증폭된 영역과 결실된 영역을 나타내는 피크사이의 거리에 의해 반영될 수 있다. 거리가 클수록, 시료 속의 종양-기원한 DNA의 분획 농도는 커질 수 있다. 시료 속의 종양-기원한 DNA의 분획 농도는 다음 식을 사용하여, F_{GR} 로 나타낸, 당해 계놈 표시 시도로 측정할 수 있다:

[0204] $F_{GR} = R_{우측} - R_{좌측}$ (여기서, $R_{우측}$ 은 우측 피크의 R 값이고 $R_{좌측}$ 은 좌측 피크의 R 값이다). 최대 차이는 1일 수 있으며, 이는 100%에 상응한다. HCC 환자로부터 수득된 종양 시료 속에서 종양-기원한 DNA의 분획 농도는 66%로 평가되며, 여기서 $R_{우측}$ 및 $R_{좌측}$ 의 값은 각각 1.376 및 0.712이다.

[0205] 당해 결과를 입증하기 위하여, 계놈전체 응집된 대립형질 손실(GAAL) 분석을 사용하는 다른 방법을 또한 사용하여 종양 DNA의 비율의 분획 농도를 독립적으로 결정하였다(참조: Chan KC 등 Clin Chem 2013;59:211-24). 표 3은 계놈 표시(F_{GR}) 및 GAAL(F_{GAAL}) 시도를 사용하여 4명의 HCC 환자의 조양 조직 속에서 종양-기원한 DNA의 분획 농도를 나타낸다. 이들 2개의 상이한 시도에 의해 측정된 값은 서로 잘 일치한다.

표 3

[0206] GAAL 및 계놈 표시(GR)에 의해 측정된 분획 농도

HCC 환자	F_{GAAL}	F_{GR}
1	60.0%	66.5%
2	60.0%	61.4%
3	58.0%	58.9%
4	45.7%	42.2%

[0207] 2. 종양 농도가 낮은 시료 속에서의 평가

[0208] 상기 분석은, 시료 DNA 중 50% 초과이 종양-기원하는 경우, 즉, 종양 DNA가 대다수 비율인 경우 종양 DNA의 분획 농도를 결정하기 위해 사용될 수 있음을 나타내었다. 앞서의 분석에서, 본 발명자는, 당해 방법을 또한 종양 유래 DNA가 적은 비(즉, 50% 미만)를 나타내는 시료에 적용할 수 있음을 나타내었다. 적은 비의 종양 DNA를 함유할 수 있는 시료는 혈액, 혈장, 혈청, 뇨, 활막액, 뇌척수액, 눈물, 타액, 복수액 및 암 환자의 변을 포함하나, 이에 한정되지 않는다. 일부 시료에서, 종양-기원한 DNA의 분획 농도는 49%, 40%, 30%, 20%, 10%, 5%, 2%, 1%, 0.5%, 0.1% 이하일 수 있다.

[0209] 이러한 시료의 경우, 증폭 및 결실을 지닌 영역을 나타내는 서열 관독물 밀도의 피크는 위에서 나열한 바와 같이 종양 유래 DNA의 비교적 높은 농도를 함유하는 시료에서와 같이 명백하지 않을 수 있다. 하나의 구현예에서, 암 세포 내 염색체 비정상성을 지닌 영역은 암 DNA를 함유하지 않은 것으로 공지된 참조 시료와 비교함으로써 확인할 수 있다. 예를 들어, 암을 지니지 않은 피검자의 혈장을 참조물질로 사용하여 염색체 영역에 대한 서열 관독물 밀도의 규범적 범위를 측정할 수 있다. 시험한 피검자의 서열 관독물 밀도는 참조 그룹의 값과

비교할 수 있다. 하나의 구현예에서, 서열 판독물 밀도의 평균 및 표준 편차(SD)를 측정할 수 있다. 각각의 빈의 경우, 시험한 피검자의 서열 판독물 밀도를 참조 그룹의 평균과 비교하여 다음 식을 사용하여 z-점수를 결정한다:

$$z-score = \frac{(GR_{test} - \overline{GR_{ref}})}{SD_{ref}} \quad [0210]$$

, 상기 식에서, GR_{test} 는 암 환자의 서열 판독물 밀도를 나타내고; $\overline{GR_{ref}}$ 는 참조 피

검자의 평균 서열 판독물 밀도를 나타내며 SD_{ref} 는 참조 피검자에 대한 서열 판독물 밀도의 SD를 나타낸다.

[0211] < -3의 z-점수를 지닌 영역은 중앙 조직 내 결실의 존재를 나타내는 암 환자내 특수한 빈에 대한 서열 판독물 밀도의 유의적인 과소표시를 나타낸다. > 3인 z-점수를 갖는 영역은 중앙 조직 내 증폭의 존재를 나타내는 암 환자에서 특수한 빈에 대한 서열 판독물 밀도의 유의적인 과대표시를 나타낸다.

[0212] 이후에, 모든 빈의 z-점수의 분포를 구성하여 상이한 수의 카피 획득 및 손실, 예를 들면, 염색체의 1 또는 2개의 카피의 결실; 및 염색체의 1, 2, 3 및 4개의 추가의 카피를 생성하는 증폭을 지닌 영역을 확인할 수 있다. 일부 경우에, 1개 이상의 염색체 또는 염색체의 1개 이상의 영역이 포함될 수 있다.

[0213] 도 10b는 본 발명의 구현예에 따른 HCC 환자의 혈장 속에서 모든 빈에 대한 z-점수의 분포 플롯 1050을 나타낸다. 1개 카피 손실, 카피 변화 없음, 1개 카피 획득 및 2개 카피 획득을 나타내는 피크(좌측으로부터 우측으로)를 z-점수 분포에 적합하도록 한다. 이후에, 상이한 유형의 염색체 비정상성을 지닌 영역을 예를 들면, 위에서 기술한 바와 같은 정상 혼합물 모델을 사용하여 확인할 수 있다.

[0214] 시료(F) 속에서 암 DNA의 분획 농도는 1개 카피 획득 또는 1개 카피 손실을 나타내는 빈의 서열 판독물 밀도로

$$F = \frac{|GR_{test} - \overline{GR_{ref}}| \times 2}{GR_{ref}} \times 100\%$$

부터 부여될 수 있다. 특수한 빈에 대해 측정된 분획 농도는 GR_{ref} 로서 계산될 수 있다. 이는

$$F = \frac{|z-score \times SD_{ref}| \times 2}{GR_{ref}}$$

또한 GR_{ref} 로 나타낼 수 있으며, 이는 $F = |Z-점수| \times CV \times 2$ (여기서 CV는 참조 피검자의 서

$$CV = \frac{SD_{ref}}{GR_{ref}}$$

열 판독물 밀도의 측정을 위한 변이 계수이고; CV는 GR_{ref} 이다)로 다시 기술될 수 있다.

[0215] 하나의 구현예에서, 빈으로부터의 결과를 합한다. 예를 들면, 1개 카피 획득을 나타내는 빈의 z-점수를 평균내거나 수득되는 F 값을 평균낼 수 있다. 다른 시행시, F를 부여하는데 사용된 z-점수의 값은 통계적 모델로 측정되며 도 10b 및 도 11에 나타난 피크로 나타낸다. 예를 들면, 우측 피크의 z-점수를 사용하여 1개 카피 획득을 나타내는 영역에 대한 분획 농도를 측정할 수 있다.

[0216] 다른 구현예에서, < -3의 z-점수 및 >3의 z-점수를 지닌 모든 빈은, 이들 2개 유형의 염색체 비정상이 가장 일반적이므로, 단일 카피 손실 및 단일 카피 획득 각각을 지닌 영역에 기여할 수 있다. 이러한 추측은, 염색체 비정상성을 지닌 빈의 수가 비교적 작고 정상의 분포를 적합하게 하는 것이 정밀하지 않을 수 있는 경우 가장 유용하다.

[0217] 도 11은 본 발명의 구현예에 따른 HCC 환자의 혈장에 대한 z-점수의 분포 플롯 (1100)을 나타낸다. 염색체 비정상과 중첩되는 빈의 수가 비교적 작지만, < -3의 z-점수 및 >3의 z-점수를 지닌 모든 빈을 단일의 카피 손실 및 단일의 카피 획득 각각의 정상 분포에 적합하도록 하였다.

[0218] 4명의 HCC 환자의 혈장 속에서 중앙-기원한 DNA의 분획 농도를 GAAL 분석 및 당해 GR 계 시도를 사용하여 결정하였다. 당해 결과는 표 4에 나타낸다. 알 수 있는 바와 같이, 유추된 분획 표시는 GAAL 분석과 GR 분석 사이에서 잘 상호 관련되어 있다.

표 4

[0219] 염색체 비정상 분석으로 유추된 혈장 속에서 종양-기원한 DNA의 분획 농도

	혈장 속에서 종양-기원한 DNA의 분획 농도	
시료	GAAL 분석	GR 분석
경우11	4.3%	4.5%
경우13	5%	5.5%
경우23	52%	62%
경우27	7.6%	6.1%

[0220] C. 분획 농도를 결정하는 방법

[0221] 도 12는 본 발명의 구현예에 따라 세포-유리된 DNA를 포함하는 생물학적 시료 중의 종양 DNA의 분획 농도를 결정하는 방법 (1200)의 흐름도이다. 방법 (1200)은 위에서 기술한 양태를 포함하는, 각종 양태를 통해 수행할 수 있다.

[0222] 블록 (1210)에서, 하나 이상의 서열 태그는 생물학적 시료 중의 다수의 DNA 단편 각각에 대해 접수된다. 블록 (1210)은 다른 방법의 경우 본원에 기술된 바와 같이 수행될 수 있다. 예를 들어, DNA 분획의 한쪽 말단을 혈장 시료로부터 서열분석할 수 있다. 다른 구현예에서, DNA 분획의 양쪽 말단을 서열분석함으로써, 단편의 길이를 평가하도록 할 수 있다.

[0223] 블록 (1220)에서, 게놈 위치는 서열 태그를 위해 측정한다. 게놈 위치는 예를 들면, 본원에 기술된 바와 같이 서열 태그를 기준 게놈에 대해 정렬함으로써 측정할 수 있다. 단편의 말단 둘 다를 서열분석한 경우, 쌍을 이룬 태그는 규정된 거리, 예를 들면, 500 또는 1,000개 기준 미만으로 구축된 2개의 태그 사이의 거리를 지닌 쌍으로 정렬시킬 수 있다.

[0224] 블록 (1230)에서, 다수의 게놈 영역 다수에 대해, 게놈 영역 내 DNA 단편의 각각의 양은 게놈 영역 내에 게놈 위치를 가진 서열 태그로부터 측정된다. 게놈 영역은 기준 게놈 내 동일한 길이의 중첩되지 않은 빈일 수 있다. 하나의 구현예에서, 빈에 대해 정렬되는 태그의 수를 계수할 수 있다. 따라서, 각각의 빈은 상응하는 수의 정렬된 태그를 가질 수 있다. 히스토그램은, 빈이 특정 수의 정렬된 태그를 갖는 빈도를 나타내도록 컴퓨터처리할 수 있다. 방법 (1200)은 각각 동일한 길이(예를 들면, 1Mb 빈)를 갖는 게놈 영역에 대해 수행할 수 있으며, 여기서 당해 영역은 중첩되지 않는다. 다른 구현예에서, 상이한 길이를 사용할 수 있으며, 이는 계수될 수 있고, 당해 영역은 중첩될 수 있다.

[0225] 블록 (1240)에서, 각각의 양을 정규화하여 각각의 밀도를 수득한다. 하나의 구현예에서, 각각의 양을 정규화하여 각각의 밀도를 수득하는 것은 동일한 총 수의 정렬된 기준 태그를 사용하여 각각의 밀도 및 기준 밀도를 결정하는 것을 포함한다. 다른 구현예에서, 각각의 양은 정렬된 기준 태그의 총 수로 나눌 수 있다.

[0226] 블록 (1250)에서, 각각의 밀도를 기준 밀도와 비교하여, 게놈 영역이 1개 카피 손실 또는 1개 카피 획득을 나타내는지 확인한다. 하나의 구현예에서, 차이를 각각의 밀도와 기준 밀도 사이에서(예를 들면, z-점수를 결정하는 부분으로서) 컴퓨터처리하여 컷오프 값을 비교한다. 각종 구현예에서, 기준 밀도는 건강한 세포의시료로부터(예를 들면, 연막으로부터) 또는 각각의 양 자체로부터(예를 들면, 대부분의 영역이 손실 또는 획득을 나타내지 않는다는 가정하에서, 중간 또는 평균값을 고려함으로써) 수득할 수 있다.

[0227] 블록 (1260)에서, 제1 밀도는 1개 카피 획득을 나타내는 것으로서 확인된 하나 이상의 각각의 밀도로부터 또는 1개 카피 손실을 나타내는 것으로 확인된 하나 이상의 각각의 밀도로부터 계산한다. 제1 밀도는 바로 하나의 게놈 영역에 상응할 수 있거나, 다수의 게놈 영역의 밀도로부터 측정할 수 있다. 예를 들면, 제1 밀도는 1개 카피 손실을 갖는 각각의 밀도로부터 컴퓨터처리할 수 있다. 각각의 밀도는 종양 농도를 제공하는 종양 내 영역의 결실로부터 생성되는 밀도 차이의 양의 척도를 제공한다. 유사하게, 제1 밀도가 1개 카피 획득을 갖는 각각의 밀도로부터 기원한 경우, 종양 내 영역의 2배로부터 생성된 밀도 차이의 양의 척도를 수득할 수 있다. 상기 단락은, 다수 영역의 밀도를 사용하여 제1 밀도에 대해 사용될 평균 밀도를 결정하는 방법의 각종 예를 기술한다.

[0228] 블록 (1270)에서, 분획 농도는 제1 밀도를 다른 밀도와 비교하여 차이를 수득함으로써 계산한다. 차이는 기준 밀도로 정규화되며, 이는 블록 (1270)에서 수행될 수 있다. 예를 들면, 상기 차이는 상기 차이를 기준 밀도로 나눔으로써 기준 밀도로 정규화할 수 있다. 다른 구현예에서, 차이는 앞서의 블록에서 정규화될 수 있다.

[0229] 하나의 시행에서, 다른 밀도는 예를 들면, 상기 단락 2에서와 같은 기준 밀도이다. 따라서, 분획 농도를 계산

하는 것은 상기 차이에 2를 곱함을 포함할 수 있다. 다른 시행에서, 다른 밀도는 예를 들면, 상기 단락 1에 기술된 바와 같이, 1개 카피 손실(여기서 제1 밀도는 1개 카피 획득을 나타내는 것으로 확인된 각각의 밀도를 사용하여 계산된다)을 나타내는 것으로 확인된 각각의 밀도로부터 계산된 제2 밀도이다. 이 경우, 정규화된 차이는 제1 밀도의 제1 비(예를 들면, $R_{우측}$) 및 기준 밀도를 계산하고 제2 밀도와 기준 밀도의 제2 비($R_{좌측}$)를 계산함으로써 측정할 수 있으며, 여기서 상기 차이는 제1 비와 제2 비 사이에 있다. 위에서 기술된 바와 같이, 1개 카피 손실 또는 1개 카피 획득을 나타내는 게놈 영역의 확인은 각각의 밀도의 히스토그램의 분포 곡선에 피크를 적합하게 함으로써 수행할 수 있다.

[0230] 요약하면, 구현에는 상이한 염색체 영역 내 혈장 DNA의 게놈 표시를 분석하여 염색체 영역이 종양 조직 내에서 증폭되거나 결실되는지의 여부 및, 당해 영역이 증폭되는지 또는 결실되는지를 동시에 측정할 수 있으며, 이의 게놈 표시를 사용하여 종양 유래 DNA의 분획 농도를 유추할 수 있다. 일부 시행은 정상의 혼합물 모델을 사용하여 상이한 영역의 게놈 표시의 전체 분포를 분석함으로써 상이한 유형의 비정상, 즉 1, 2, 3 또는 4개의 카피의 획득 및 1 또는 2개의 카피의 손실과 관련된 게놈 표시를 측정한다.

[0231] 구현에는 다른 방법, 예를 들면, 게놈전체 응집된 대립형질 손실(GAAL) 시도(미국 특허원 제13/308473호; Chan KC 등 Clin Chem 2013;59:211-24) 및 종양 관련된 단일 뉴클레오타이드 돌연변이의 분석(참조: Forshew T 등 Sci Transl Med. 2012;4: 136ra68)보다 몇가지 장점을 가진다. 염색체 비정상을 가진 영역에 대한 모든 서열 판독물 맵핑을 사용하여 영역의 서열 판독물 밀도를 결정하며, 따라서 종양 DNA의 분획 농도에 관한 정보를 제공한다. 다른 한편, GAAL 분석에서, 개인에서 이형접합성이고 염색체 획득 또는 손실을 가진 염색체 영역 내에 위치하는 단일 뉴클레오타이드를 포함하는 서열 판독물만이 정보를 제공할 수 있다. 유사하게, 암-관련된 돌연변이의 분석의 경우, 돌연변이를 포함하는 서열 판독물만이 종양 DNA 농도의 유추를 위해 유용할 수 있다. 따라서, 구현에는, 비교적 더 적은 서열분석 판독물이 다른 시도와 비교하는 경우 종양 유래 DNA의 분획 농도의 평가시 동일한 정도의 정확도를 달성하기 위해 요구될 수 있으므로 서열분석 데이터의 보다 비용-효과적인 사용을 허용할 수 있다.

[0232] X. 대안적인 방법

[0233] 실제 돌연변이인 것으로 유전자좌를 확인하기 위한 기준으로서 특수한 돌연변이가 서열 태그 상에 관찰되는 횟수를 사용(함으로써 양성 예측값을 조절)하는 것과는 별개로, 컷오프 값을 사용하는 것 대신 또는 이외에 다른 기술을 사용하여 양성 돌연변이를 확인하는데 있어 더 큰 예측된 값을 제공할 수 있다. 예를 들면, 서열 분석 데이터를 예를 들면, 서열분석된 뉴클레오타이드의 품질 점수를 고려함으로써 프로세싱하는 경우 상이한 전략의 생물정보 필터를 사용할 수 있다. 하나의 구현예에서, DNA 서열분석기 및, 상이한 서열분석 오차 프로파일의 지닌 서열분석 화학을 사용할 수 있다. 보다 낮은 서열분석 오차율을 지닌 서열분석기 및 화학은 보다 높은 양성 예측값을 제공할 수 있다. 또한 동일한 DNA 단편의 반복된 서열분석을 사용하여 서열분석 정밀성을 증가시킬 수 있다. 한가지 가능한 전략은 업자(Pacific Biosciences)의 원형 컨센스 서열분석 전략이다.

[0234] 다른 구현예에서, 서열분석된 단편상의 크기 정보를 데이터의 해석으로 도입할 수 있다. 종양 유래 DNA는 혈장 속의 종양 기원하지 않은 DNA보다 더 짧으므로(참조: 미국 특허원 제13/308,473호), 잠재적인 종양 기원한 돌연변이를 함유하는 보다 짧은 혈장 DNA 단편의 양성 예측값은 보다 긴 혈장 DNA 단편의 것보다 더 높을 것이다. 크기 데이터는, 혈장 DNA의 쌍을 이룬-말단 서열분석을 수행하는 경우 용이하게 이용가능할 것이다. 대안으로서, 긴 판독물 길이를 지닌 DNA 서열을 사용함으로써, 완전한 길이의 혈장 DNA 단편을 수득할 수 있다. 또한 DNA 서열분석 전에 혈장 DNA 시료의 크기 분획화를 수행할 수 있다. 크기 분획화를 위해 사용할 수 있는 방법의 예는 겔 전기영동, 미세유체 시도의 사용(예를 들면, Caliper LabChip XT 시스템) 및 크기 배제 스핀 컬럼을 포함한다.

[0235] 여전히 다른 구현예에서, 비 혈액암을 지닌 환자에서 혈장 내 종양 관련된 돌연변이의 분획 농도는, 혈장 내 보다 짧은 DNA 단편에 초점을 맞추는 경우 증가하는 것으로 예측될 수 있다. 하나의 시행에서, 2개 이상의 상이한 크기 분포의 DNA 단편에 있어서 혈장 내 종양 관련된 돌연변이의 분획 농도를 비교할 수 있다. 비 혈액암을 지닌 환자는 보다 큰 분획과 비교하는 경우 보다 짧은 분획내 종양 관련된 돌연변이의 보다 큰 분획 농도를 지닐 것이다.

[0236] 일부 구현예에서, 동일한 혈액 시료의 2개 이상의 분취량, 또는 동일한 경우 또는 상이한 경우에 고려된 2개 이상의 혈액 시료로부터 서열 분석 결과를 합할 수 있다. 1개 이상의 분취량 또는 시료에서 관찰된 잠재적인 돌연변이는 종양 관련된 돌연변이의 보다 높은 양성 예측값을 가질 수 있다. 양성 예측값은 이러한 돌연변이를 나타내는 시료의 수를 증가시킬 수 있다. 상이한 시점에서 취한 혈장 시료 속에 존재하는 잠재적인 돌연

변이는 잠재적 돌연변이로 고려될 수 있다.

[0237] **XI. 실시예**

[0238] 다음은 예시적인 기술 및 데이터이며, 이는 본 발명의 구현예를 제한하는 것으로 고려되어서는 안된다.

[0239] **A. 물질 및 방법**

[0240] 시료 수집과 관련하여, 간세포 암종(HCC) 환자, 만성 B형 간염의 매개체, 및 동시발생한 유방 및 난소암을 지닌 환자를 모집하였다. 모든 HCC 환자는 바르셀로나 임상 간암 AI기 질병을 가졌다. 모든 참여자로부터의 말초 혈액 시료를 EDTA를 함유하는 튜브내로 수집하였다. HCC 환자의 종양 조직은 이들의 암 절개 수술 동안 수득하였다.

[0241] 말초 혈액 시료를 4℃에서 1,600g로 10분 동안 원심분리하였다. 혈장 부위를 4℃에서 16,000 g로 10분 동안 재원심분리한 후 80℃에서 저장하였다. 4.8mL의 혈장으로부터의 무세포 DNA 분자를 QIAamp DSP DNABlood Mini Kit(제조원: Qiagen)의 혈액 및 체액 프로토콜에 따라 추출하였다. 혈장 DNA를 SpeedVac 농축기(Savant DNA 120; 제조원: Thermo Scientific)를 사용하여 DNA-서열분석 라이브러리의 후속적인 제조를 위해 경우당 40 µl의 최종 용적으로 농축시켰다.

[0242] 게놈 DNA를 환자의 연막 시료로부터 QIAamp DSP DNA Blood Mini Kit의 혈액 및 체액 프로토콜에 따라 추출하였다. DNA를 QIAamp DNA Mini Kit(제조원: Qiagen)을 사용하여 종양 조직으로부터 추출하였다.

[0243] 게놈 DNA 시료의 서열분석 라이브러리를 Paired-End Sample Preparation Kit(제조원: Illumina)를 사용하여 제조업자의 지시에 따라 작성하였다. 요약하면, 1 내지 5µg의 게놈 DNA를 우선 Covaris S220 Focused-ultrasonicator를 사용하여 200-bp 단편으로 전단시켰다. 이후에, DNA 분자를 T4 DNA 폴리머라제 및 클레노우 폴리머라제(Klenow polymerase)로 말단-보수하고; 이후에 T4 폴리뉴클레오타이드 키나제를 사용하여 5' 말단을 포스포릴화하였다. 3' 오버행(overhang)을 3' 내지 5' 엑소뉴클레아제-결핍성 클레노우 단편으로 생성시켰다. Illumina 어댑터 올리고뉴클레오타이드를 점성 말단에 연결시켰다. 어댑터 연결된 DNA를 12주기 PCR로 농축시켰다. 혈장 DNA 분자는 짧은 단편이고 혈장 시료 속의 전체 DNA의 양은 비교적 작으므로, 본 발명자는 단편화 단계를 생략하고 혈장 시료로부터 DNA 라이브러리를 작성시 15주기 PCR을 사용하였다.

[0244] Agilent 2100 Bioanalyzer(제조원: Agilent Technologies)을 사용하여 어댑터-연결된 DNA 라이브러리의 품질 및 양을 점검하였다. 이후에, DNA 라이브러리를 KAPA Library Quantification Kit(제조원: Kapa Biosystems)에 의해 제조업자의 지시에 따라 결정하였다. DNA 라이브러리를 희석시키고 쌍을 이룬 말단 서열분석 유동 셀에 하이브리드화하였다. DNA 클러스터(cluster)를 cBot 클러스터 생성 시스템(제조원: Illumina) 상에서 TruSeq PE Cluster Generation Kit v2(제조원: Illumina)을 사용한 후 HiSeq 2000 시스템(제조원: Illumina) 상에서 TruSeq SBS Kit v2 (제조원: Illumina)를 사용한 51_2 주기 또는 76_2 주기의 서열분석에 의해 생성시켰다.

[0245] 쌍을 이룬 말단 서열분석 데이터를 쌍을 이룬 말단 방식으로 짧은 올리고뉴클레오타이드 정렬 프로그램 2(Short Oligonucleotide Alignment Program 2: SOAP2)를 사용하여 분석하였다. 각각의 쌍을 이룬 말단 관독물에 대해, 각각의 말단으로부터 50bp 또는 75bp를 비-반복물-차폐된 참조 사람 게놈(hg1 8)에 정렬하였다. 2개 이하의 뉴클레오타이드 미스매치(mismatch)를 각각의 말단의 정렬을 위해 허용하였다. 2개 말단에 대한 이들 잠재적인 정렬의 게놈성 조화는 이후에 2개의 말단이 동일한 염색체에 정확한 배향으로 정렬되어, 삽입체 크기를 600bp 이하로 확장하도록 하며, 참조 사람 게놈 내 단일 위치에 대해 맵핑하도록 할 수 있다. 중복된 관독물은 쌍을 이룬 말단 관독물로 정의되었으며, 여기서 삽입체 DNA 분자는 사람 게놈 내 동일한 출발 및 말단 위치를 나타내었으며; 중복 관독물은 앞서 기술한 바와 같이 제거하였다(참조: Lo 등 Sci Transl Med 2010; 2: 61ra91).

[0246] 일부 구현예에서, 쌍을 이룬 종양 및 구성적 DNA 시료를 서열분석하여 종양 관련된 단일의 뉴클레오타이드 변이체(SNV)를 확인하였다. 일부 시행에서, 본 발명자들은 구성적 DNA(당해 실시예에서는 연막 DNA임)내 동형접합성 부위에서 발생하는 SNV에 초점을 맞추었다. 원칙적으로, 종양 조직의 서열분석 데이터 내에서 검출되지만 구성적 DNA 내에서는 부재한 어떠한 뉴클레오타이드 변이도 잠재적 돌연변이(즉, SNV)일 수 있었다. 그러나, 서열분석 오차(0.1% 내지 0.3%)로 인하여, 종양 조직의 서열분석 데이터 내 어떠한 뉴클레오타이드 변화의 단일 발생이 종양-관련 SNV로 고려된 경우 수백만의 거짓 양성체가 게놈 내에서 확인될 수 있다. 거짓 양성체의 수를 감소시키는 한가지 방법은 종양 관련된 SNV를 요청할 수 있기 전에 종양 조직 내 서열분석 데이터 내 동일한 뉴클

레오타이드 변화의 다수발생을 관측하는 기준을 세우는 것일 수 있다.

- [0247] 서열분석 오차의 발생은 확률적 과정이므로, 서열분석 오차로 인한 거짓 양성(+)의 수는 종양 관련된 SNV로서 관찰된 SNV가 자격을 갖추는데 요구되는 발생의 수의 증가와 함께 기하급수적으로 감소할 수 있다. 한편, 거짓 양성(+)의 수는 서열분석 깊이가 증가함에 따라 증가할 수 있다. 이들 관계는 푸아송 및 이항 분포 함수로 예측할 수 있었다. 구현에는 관찰된 SNV를 종양 관련된 것으로 자격을 부여하기 위한 발생의 역학적 컷오프를 측정할 수 있다. 구현에는 종양 서열분석 데이터 내 특수 뉴클레오타이드의 실제 포함, 서열분석 오차율, 허용된 최대 거짓-양성율, 및 돌연변이 검출을 위한 요구되는 민감성을 고려할 수 있다.
- [0248] 일부 예에서, 본 발명자들은 매우 엄격한 기준을 설정하여 거짓 양성을 감소시켰다. 예를 들면, 돌연변이가 구성적 DNA 서열분석내에 완전히 부재함이 요구될 수 있으며, 특수 뉴클레오타이드 위치에 대한 서열분석 깊이는 20배이어야 했다. 일부 시행시, 발생의 컷오프는 10^{-7} 미만의 거짓-양성 검출률을 달성하였다. 일부 실시예에서, 본 발명자들은 동원체성, 텔로머성(telomeric), 및 저-복합성 영역 내에 있던 SNV를 필터링 제거하여 정렬 결함으로 인한 거짓 양성을 최소화시켰다. 또한, dbSNP 빌드 135 데이터베이스((dbSNP build 135 database)내 공지된 SNP에 대한 추정적 SNV 매핑을 또한 제거하였다.
- [0249] B. 절제 전 및 후
- [0250] 도 13a는 본 발명의 구현예에 따른 진단 시기에 난소암 및 유방암을 지닌 환자의 혈장 속의 돌연변이 분석의 표 1300을 나타낸다. 여기서, 본 발명자는 쌍방의 난소암 및 유방암을 지닌 환자에 대한 예를 입증한다. 혈장의 서열분석 데이터를 환자의 구성적 DNA(연막)의 서열분석 결과와 비교하였다. 혈장 속에는 존재하지만 구성적 DNA속에는 존재하지 않았던 단일 뉴클레오타이드 변화를 잠재적 돌연변이로 간주하였다. 환자의 우측 및 좌측의 난소암을 2개 부위에서 각각 시료채취하고, 총 4개의 종양 시료를 제조하였다. 종양 돌연변이는 4개의 상이한 부위에서 4개의 난소 종양 조직 내에서 검출된 돌연변이였다.
- [0251] 360만개 이상의 단일 뉴클레오타이드 변화가 서열분석에 의해 적어도 1회 동안 혈장 속에서 검출되었다. 이들 변화 중에서, 단지 2,064개 만이 종양 조직 속에서 검출되어 0.06%의 양성 예측값을 제공하였다. 혈장 속에서 적어도 2회 검출되는 기준을 사용하면, 잠재적 돌연변이의 수가 18,885에 대해 99.5%까지 유의적으로 감소하였다. 종양 돌연변이의 수는 2,003으로 3%로 감소하였으며, 양성(+)의 예측 값은 11%로 증가하였다.
- [0252] 혈장 속에서 적어도 5회 검출하는 기준을 사용하여, 단지 2,572개의 잠재적 돌연변이가 검출되었고, 이들 중에서, 1,814개가 모든 종양 조직 속에서 검출된 돌연변이였으므로, 71%의 양성 예측값을 제공하였다. 발생 수(예를 들면, 2, 3, 4, 6, 7, 8, 9, 10 등)에 대한 다른 기준은 요구되는 민감성 및 양성(+)의 예측값에 따라 잠재적 돌연변이를 정의하기 위해 사용할 수 있다. 기준으로 사용되는 발생의 수가 클수록, 민감성에 있어서 감소될 수 있는 양성(+)의 예측값이 더 높아진다.
- [0253] 도 13b는 본 발명의 구현예에 따른 종양 절개 후 쌍방의 난소암 및 유방암을 지닌 환자의 혈장 속에서 돌연변이의 분석의 표 1350을 나타낸다. 환자의 외과적 절개를 수행하였다. 혈액 시료를 난소 암 및 유방암의 절개 후 1일째에 취하였다. 이후에, 혈장 DNA를 서열분석하였다. 당해 실시예의 경우, 난소암으로부터의 돌연변이만이 분석되었다. 3백만개 이상의 잠재적인 돌연변이가 혈장 시료 속에서 적어도 1회 검출되었다. 그러나, 적어도 5개의 발생을 갖는 기준을 사용하여, 잠재적 돌연변이의 수는 238로 감소되었다. 진단시 취한 시료에 대한 및 5개 돌연변이의 동일한 기준을 사용하여 잠재적 돌연변이의 수를 비교하는 경우 유의적인 감소가 관찰되었다.
- [0254] 하나의 구현예에서, 혈장 속에서 검출된 단일 뉴클레오타이드 변화의 수는 암 환자의 검출, 모니터링 및 예후를 위한 매개변수로서 사용할 수 있다. 상이한 수의 발생을 기준으로 사용하여 바람직한 민감성 및 특이성을 달성할 수 있다. 종양 부하가 더 높고 따라서 열악한 예후를 지닌 환자는 혈장 속에서 관찰된 보다 높은 돌연변이 부하를 가진 것으로 예측될 것이다.
- [0255] 이러한 분석을 위해, 상이한 유형의 암에 대한 돌연변이 부하 프로파일을 확립할 수 있다. 모니터링 목적을 위해, 치료에 대해 반응하는 환자의 혈장 속의 돌연변이 부하가 감소할 수 있는지를 관찰할 수 있다. 종양이 예를 들면, 재발 동안, 재발생하는 경우, 돌연변이 부하는 증가하는 것으로 예측될 것이다. 이러한 모니터링은 환자에 대한 치료의 선택된 양식의 효능을 모니터링하여 특수 치료에 대한 내성의 출현을 검출할 수 있도록 할 수 있다.
- [0256] 혈장 DNA 서열분석 결과에서 발견할 수 있는 특이적인 돌연변이의 분석을 통해, 민감성(예를 들면, 상피 성장인자 수용체 내에서 돌연변이 및 타이로신 키나제 억제제 치료에 대한 반응) 및 특이적인 표적화된 치료(예를

들면, 결장직장암에서 KRAS 돌연변이 및 파니투무마브 및 세특시마브에 의한 치료에 대한 내성)에 대한 내성을 예측할 수 있는 표적을 또한 확인할 수 있고 치료 요법의 계획을 안내할 수 있다.

[0257] 상기 예는 쌍방의 난소 암에 대한 것이었다. 유방암의 돌연변이에서 동일한 분석을 또한 수행한 후 혈장 속에서 이들 암 유형 둘 다의 돌연변이를 추적할 수 있다. 또한 유사한 전략을 사용하여 주요 암의 돌연변이 및 이의 전이 또는 전이들을 추적할 수 있다.

[0258] 구현에는 특수한 위험 인자(예를 들면, 흡연 상태, 바이러스 상태(예를 들면, 간염 바이러스 매개체, 사람 파필로마바이러스 감염된 피검자))를 지닌 피검자 또는 명백하게 건강한 피검자에서 암의 스크리닝에 유용할 수 있다. 이러한 피검자의 혈장 속에서 관찰할 수 있는 돌연변이 부하는, 피검자가 특수한 기간 내에서 증상이 있는 암으로 발달할 위험을 제공할 수 있다. 따라서, 혈장 속에서 보다 높은 돌연변이 부하를 지닌 피검자는 보다 낮은 돌연변이 부하를 지닌 피검자보다 더 높은 위험을 갖는 것으로 예측될 수 있다. 또한, 혈장 속에서 이러한 돌연변이 부하의 일시적인 프로파일은 또한 위험의 강력한 지시인자일 수 있다. 예를 들어, 피검자가 매년 수행된 하나의 혈장 돌연변이 부하를 가지고 이러한 돌연변이 부하가 점진적으로 증가하는 경우, 당해 피검자는 예를 들면, 가슴 X 선, 초음파, 컴퓨터 단층촬영, 자기 공명 영상 또는 양전자 방사 단층촬영을 사용하여, 암에 대한 추가의 스크리닝 양상에 대해 언급할 수 있다.

[0259] C. 서열분석 혈장으로부터 돌연변이를 유추하기 위한 역학적 컷오프

[0260] 간세포 암종(HCC)을 지닌 4명의 환자 및 난소암 및 유방암을 지닌 1명의 환자를 본 연구를 위해 모집하였다. 후자의 환자의 경우, 본 발명자들은 난소암의 분석에 초점을 맞추었다. 혈액 시료를 종양의 외과적 절개 전 및 후에 각각의 환자로부터 수집하였다. 절개된 종양 조직을 또한 수집하였다. 종양 조직으로부터 추출된 DNA, 수술전 혈액 시료의 백혈구 세포 및 수술 전 및 후 혈장 시료를 HiSeq2000 서열분석 시스템(제조사: Illumina)을 사용하여 서열분석하였다. 서열분석 데이터를 참조 사람 게놈 서열(hg18)에 대해 짧은 올리고뉴클레오타이드 분석 패키지 2(SOAP2)를 사용하여 정렬하였다(참조: Li R 등 Bioinformatics 2009; 25: 1966-1967). 백혈구 세포의 DNA 서열은 각각의 연구 피검자에 대해 구성적 DNA 서열로서 간주하였다.

[0261] 당해 실시예에서, 종양-관련된 SNM은 우선 혈장 DNA 서열분석 데이터 및 CG로부터 종양 조직에 대한 참조없이 유추하였다. 이후에, 혈장으로부터 유추된 결과를 종양 조직으로부터 생성된 서열분석 데이터와(금표준물로서) 비교하여 유추된 결과의 정확성을 확인하였다. 이와 관련하여, 금표준물은 종양 조직 및 구성적 서열로부터의 서열분석 데이터를 비교하여 종양 조직 내에서 돌연변이를 산출함으로써 달성하였다. 당해 분석에서, 본 발명자들은, 연구된 피검자의 구성적 DNA가 동형접합성인 뉴클레오타이드 위치에 초점을 맞추었다.

[0262] 1. 표적화되지 않는 전체 게놈 분석

[0263] 각각의 환자의 백혈구 세포, 종양 조직 및 혈장 DNA에 대한 서열분석 깊이를 표 5에 나타낸다.

표 5

4명의 HCC 경우에 대한 상이한 시료의 중간 서열분석 깊이

경우	중간 서열분석 깊이(배)			
	백혈구 세포	종양 조직	수술전 혈장	수술후 혈장
HCC1	39	29	23	24
HCC2	39	29	25	28
HCC3	46	33	18	21
HCC4	46	27	20	23
난소암 환자	44	53	37	28

[0264]

[0265] 표 1에 나타난 바와 같은 혈장 돌연변이(r)을 정의하기 위한 최소 발생에 대한 역학적 컷오프를 각각의 환자의 혈장 속에서 돌연변이를 확인하는데 사용한다. 각각의 유전자좌의 서열분석 깊이가 변할 수 있으므로, 컷오프도 변할 수 있으며, 이는 유전자좌에 대한 관독물의 전체 수에 있어서 컷오프의 의존성을 효과적으로 제공한다. 예를 들면, 중간 깊이가 50 미만이라도 해도(표 5), 개개 유전자좌의 서열분석 깊이는 많이 변할 수 있고 >100 배를 포함할 수 있다.

[0266] 서열분석 오차 외에도, 다른 오차 공급원은 정렬 오차일 수 있다. 이러한 유형의 오차를 최소화하기 위해, 돌

연변이를 수반하는 서열 판독물을 Bowtie 정렬 프로그램(참조: Langmead B 등 Genome Biol 2009, 10:R25)을 사용하여 기준 게놈에 대해 재정렬하였다. SOAP2 및 Bowtie에 의한 기준 게놈의 유일한 위치에 정렬될 수 있는 판독물만을 혈장 돌연변이에 대한 하부 분석에 사용하였다. 상이한 알고리즘을 기준으로 한 정렬 소프트웨어 패키지의 다른 조합을 또한 사용할 수 있다.

[0267] 실제 서열분석 데이터 내 서열분석 및 정렬 오차를 추가로 최소화하기 위하여, 본 발명자는 서열 판독물내 단일의 뉴클레오타이드 변이를 나타낸 뉴클레오타이드 위치를 호출하기 위한 2개의 추가의 필터링 알고리즘을 적용하였다: (1) 돌연변이를 수반하는 서열 판독물의 $\geq 70\%$ 를 맵핑 품질이 $\geq Q20$ (즉, 잘못정렬 확률 $< 1\%$)인 Bowtie를 사용하여 동일한 게놈 배위에 재정렬할 수 있었다; (2) 돌연변이를 수반하는 서열 판독물의 $\geq 70\%$ 는 서열 판독물의 말단 둘 다(즉, 5' 및 3' 말단)의 5bp 내에 있지 않았다. 서열분석 오차가 서열 판독물의 말단 둘 다에서 보다 우세하였으므로, 이러한 이러한 필터링 규칙이 확립되었다.

[0268] 본 발명자들은 또한 종양 게놈의 사전 인식없이 종양의 유추에 영향을 미치는 인자를 시험하였다. 하나의 이러한 매개변수는 혈장 속에서 종양-기원한 DNA의 분획 농도이었다. 당해 매개변수는 다른 금 표준 매개변수로서 고려될 수 있었으며 GAAL을 사용하는 종양 게놈의 사전 지식을 사용하여 참조 목적으로 유추하였다.

[0269] 표 6은 처리 전 및 처리에 걸쳐 혈장 내에서 검출된 뉴클레오타이드 변이를 나타낸다. HCC1의 경우, 종양 게놈의 사전 지식없이, 총 961개의 단일 뉴클레오타이드 변이가 검출되었다. 혈장 속에서 검출된 이들 뉴클레오타이드 변이 중에서, 828개가 암-관련 돌연변이였다. HCC의 외과적 절개 후, 뉴클레오타이드 변이의 총 수는 43으로 감소하였고 이들 중 어느 것도 암 관련된 돌연변이가 아니었다.

[0270] 참조 목적을 위해, 수술 전 혈장 시료 속에서 종양-기원한 DNA의 분획 농도는 53%이었으며 종양 게놈의 사전 지식으로 유추되었다. HCC2, HCC3 및 HCC4의 경우, 종양 게놈의 사전 지식없이, 혈장 속의 단일 뉴클레오타이드 변이의 수는 수술전 혈장 시료에 대해 27 내지 32의 범위인 것으로 유추되었다. 이들 결과는 대략 20배의 서열 분석 깊이를 사용하여, 매우 낮은 퍼센트의 암-관련 돌연변이가 혈장 속에서 검출될 수 있었고 혈장 속에서 검출된 서열 변이의 대부분이 서열분석 오차에 기인하였다는 수학적 예측과 양립성이다. 종양 절개 후, 검출된 서열 변이의 수에 있어서 유의적인 변화는 없었다. 참조 목적을 위해, 혈장 속에서 종양-기원한 DNA의 분획 농도는 2.1% 내지 5%의 범위로 유추되었으며 종양 게놈의 사전 지식으로 유추되었다.

표 6

[0271] 혈장 속에서 검출된 뉴클레오타이드 변이

	수술 전 혈장			수술 후 혈장		
	종양-기원한 DNA의 분획 농도	단일 뉴클레오타이드 변이의 총 수	확인된 암-관련 돌연변이의 수	종양-기원한 DNA의 분획 농도	단일 뉴클레오타이드 변이의 총 수	확인된 암-관련 돌연변이의 수
HCC1	53%	961	828	0.4%	43	0
HCC2	5%	32	0	0.6%	49	0
HCC3	2.1%	29	0	0.2%	32	0
HCC4	2.6%	27	0	1.3%	35	1
난소	46%	1718	1502	0.2%	2	0

[0272] 2. 엑손의 표적 농축

[0273] 위에서 논의한 바와 같이, 목적인 영역에 대한 서열분석 깊이의 증가는 혈장 속에서 암-관련된 돌연변이를 확인하기 위한 민감성 및 특이성 둘 다를 증가시킬 수 있으므로, 암 환자와 비-암 피검자 사이의 구별력을 증가시킨다. 전체 게놈에 대한 서열분석 깊이의 증가는 여전히 매우 비용소모적인 반면, 한가지 대안은 서열분석에 대한 특정의 영역에 대해 농축시키는 것이다. 하나의 구현예에서, 선택된 엑손 또는 실제로 전체 진유전체는 서열분석을 위해 표적-농축될 수 있다. 이러한 시도는 서열 판독물의 전체 양을 증가시키지 않고 표적 영역의 서열분석 깊이를 유의적으로 증가시킬 수 있다.

[0274] HCC 환자 및 난소(및 유방) 암을 지닌 환자의 혈장 DNA의 서열분석 라이브러리는 진유전체의 표적 농축을 위한 Agilent SureSelect All Exon kit를 사용하여 포획하였다. 엑손-농축된 서열분석 라이브러리를 이후에 HiSeq 2000 서열분석 시스템을 사용하여 서열분석하였다. 서열 판독물을 사람 기준 게놈(hg18)에 대해 정렬하였다. 정렬 후, 엑손에 대해 유일하게 맵핑한 서열 판독물을 단일의 뉴클레오타이드 변이에 대해 분석하였다. 진유전체 포획 분석을 위한 혈장 내 단일 뉴클레오타이드 변이의 확인을 위해, 표 2에 나타난 역학적 컷오프 값을 사

용한다.

[0275] 도 14a는 HCC1에 대한 혈장 DNA내 단일 뉴클레오타이드 변이의 검출을 나타내는 표 1400이다. 종양 계놈의 사전 지식없이, 본 발명자는 표적화된 서열분석 데이터로부터 혈장 내 총 57개의 단일 뉴클레오타이드 변이를 유추하였다. 종양 조직으로부터 수득된 서열분석 데이터로부터의 후속적인 확인시, 55개는 실제 종양 관련된 돌연변이인 것으로 밝혀졌다. 앞서 논의한 바와 같이, 수술전 혈장 내 종양 유래 DNA의 분획 농도는 53%였다. 종양 절개 후, 혈장으로부터 수득된 표적화된 서열분석 데이터 내에서 단일의 뉴클레오타이드 변이는 검출되지 않았다. 이들 결과는, 혈장 내 단일의 뉴클레오타이드 변이의 수의 정량적 분석이 암 환자의 질병 진행을 모니터링하는데 사용될 수 있음을 나타낸다.

[0276] 도 14b는 HCC2에 대한 혈장 DNA내 단일 뉴클레오타이드 변이의 검출을 나타내는 표 1450이다. 종양 계놈의 사전 지식없이, 본 발명자는 혈장의 표적화된 서열분석 데이터로부터 총 18개의 단일 뉴클레오타이드 변이를 유추하였다. 이들 변이중 모두가 종양 조직 내에서 발견되었다. 앞서 논의한 바와 같이, 수술 전 혈장 내 종양 유래 DNA의 분획 농도는 5%였다. 종양 절개 후, 단일의 뉴클레오타이드 변이는 혈장 속에서 검출되지 않았다. 혈장 속에서 종양-기원한 DNA의 보다 높은 분획 농도를 갖는 HCC1과 비교하여, 보다 적은 단일의 뉴클레오타이드 변이가 HCC2를 포함하는 경우의 혈장 속에서 검출되었다. 이들 결과는, 혈장 내 종양 유래 DNA의 농도가 종양 부하와 긍정적으로 상호 관련되어 있음이 밝혀져 있으므로(참조: Chan KC 등 Clin Chem 2005;51 :2192-5) 혈장 속의 단일 뉴클레오타이드 변이의 수를 매개변수로 사용하여 혈장 속에서 종양-기원한 DNA의 분획 농도, 및 따라서 환자내 종양 부하를 반영할 수 있음을 제안한다.

[0277] 도 15a는 HCC3에 대한 혈장 DNA내 단일의 뉴클레오타이드 변이의 검출을 나타내는 표 1500이다. 종양 계놈의 사전 지식없이, 본 발명자는 표적화된 서열분석 데이터로부터 수술 전 및 수술 후 둘 다의 혈장 시료 속에서 어떠한 단일의 뉴클레오타이드 변이를 관찰하지 않았다. 이는 당해 환자에서 혈장 내 종양-기원한 DNA의 비교적 낮은 분획 농도(2.1%)에 기인할 수 있다. 서열분석 깊이에 있어서 추가의 증가는 종양 유래 DNA의 낮은 분획 농도를 지닌 경우에 암 관련된 돌연변이를 검출하기 위한 민감성을 증진시키는 것으로 예측된다.

[0278] 도 15b는 HCC4에 대한 혈장 DNA내 단일의 뉴클레오타이드 변이의 검출을 나타내는 표 1550이다. 종양 계놈의 사전 지식없이, 본 발명자는 혈장의 표적화된 서열분석 데이터로부터 총 3개의 단일 뉴클레오타이드 변이를 유추하였다. 이들 돌연변이중 모두는 종양 조직에서 발견되었다. 혈장 속에서 종양 유래 DNA의 보다 높은 분획 농도를 갖는 HCC1 및 HCC2와 비교하여, 보다 적은 단일의 뉴클레오타이드 변이가 2.6%의 혈장 중 분획 종양 DNA를 가졌던 HCC4 경우의 혈장에서 검출되었다. 이들 결과는, 혈장 내 단일 뉴클레오타이드 변이의 수를 매개변수로 사용하여 혈장 내 종양 유래 DNA의 분획 농도 및 환자내 종양 부하를 반영할 수 있음을 제안한다.

[0279] 도 16은 난소(및 유방) 암 환자에 대한 혈장 DNA내 단일 뉴클레오타이드 변이의 검출을 나타내는 표 1600이다. 종양 계놈의 사전 지식없이, 본 발명자는 혈장의 표적화된 서열분석 데이터로부터 총 64개의 단일의 뉴클레오타이드 변이를 유추하였다. 이들 중 59개가 난소 종양 조직에서 발견되었다. 혈장 속에서 난소 종양 유래 DNA의 추정된 분획 농도는 46%였다. 단일 뉴클레오타이드 변이의 전체 수에 있어서의 유의적인 감소는 난소 암의 절개 후 혈장 속에서 검출되었다.

[0280] SureSelect 표적 농축 시스템(제조원: Agilent)의 사용 외에도, 본 발명자는 또한 서열분석용 엑손으로부터 서열을 농축시키기 위한 Nimblegen SeqCap EZ Exome+UTR 표적 농축 시스템(제조원: Roche)을 사용하였다. Nimblegen SeqCap 시스템은 계놈의 엑손 영역 및 또한 5' 및 3' 해독되지 않은 영역을 포함한다. 4명의 HCC 환자중 치료 전 혈장 시료, 2명의 건강한 대조군 피검자 및 암이 없는 2명의 만성 B형 간염 매개체를 분석하였다(표 7). 다른 구현예에서, 용액 상 또는 고체 상 하이브리드화를 사용하는 것을 포함하나 이에 한정되지 않는 다른 표적 농축 시스템을 사용할 수 있다.

표 7

[0281] 서열 포획용 Nimblegen SeqCap EZ Exome+UTR 표적 농축 시스템을 사용한 4명의 HCC 환자(HCC1-4)에 대한 진유 전체 서열분석 결과. HCC3의 처리 전 혈장의 서열분석 결과는 PCR-중복된 관독물의 보다 높은 퍼센트로 인하여 아-최적이었다.

처리 전 혈장				처리 후 혈장	
GAAL 분석에 의한 혈장 내 종양 유래 DNA의 분획 농도	역학적 컷오프를 충족하는 혈장 속에서 검출된 서열 변이의 수	상응하는 종양 조직 내에서 검출된 돌연변이와 중첩되는 서열 변이의 수	상응하는 종양 조직 내에서 검출된 돌연변이와 중첩되는 서열 변이의 수	역학적 컷오프를 충족하는 혈장 속에서 검출된 서열 변이의 수	상응하는 종양 조직 내에서 검출된 돌연변이와 중첩되는 서열 변이의 수

HCC1	53%	69	64	1	1
HCC2	5%	51	47	3	0
HCC3	2.1%	0	0	1	0
HCC4	2.6%	8	7	0	0

[0282] 2명의 만성 B형 간염 매개체 및 2명의 건강한 대조군 피검자에서, 역학적 컷오프 기준을 충족시킨 1개 이하의 단일 뉴클레오타이드 변이가 검출되었다(표 8). 4명의 HCC 환자 중 3명에서, 역학적 컷오프 요건을 충족시킨 혈장 내에서 검출된 서열 변이의 수는 적어도 8이었다. HCC3에서, 역학적 컷오프를 충족시킨 SNV는 검출되지 않았다. 당해 시료에서, 보다 낮은 수의 중복되지 않은 서열분석된 판독물을 초래하는 서열분석된 판독물내 높은 비율의 PCR 중복된 판독물이 존재하였다. 혈장 속에서 검출가능한 SNV의 표시된 감소는 종양의 외과적 절개 후 관찰되었다.

표 8

[0283] 서열 포획용의 Nimblegen SeqCap EZ Exome+UTR 표적 농축 시스템을 사용하여 2명의 만성 B형 간염 매개체(HBV1 및 HBV2) 및 2명의 건강한 대조군 피검자(Ctrl1 및 Ctrl2)에 대한 진염색체 서열분석 결과

	역학적 컷오프를 충족시키는 혈장 속에서 검출된 서열 변이의 수
HBV1	0
HBV2	1
Ctrl1	1
Ctrl2	1

[0284] XII. 종양 이종성

[0285] 생물학적 시료(예를 들면, 혈장/혈청) 속에서 단일 뉴클레오타이드 돌연변이의 정량화는 또한 종양 내 및 종양 간 이종성 둘 다의 종양 이종성 분석에 유용하다. 종양 내 이종성은 동일한 종양 내 종양 세포의 다수 클론의 존재에 관한 것이다. 종양간 이종성은 동일한 조직학적 유형의 2개 이상의 종양에 대한 종양 세포의 다수의 클론의 존재에 관한 것이지만, 상이한 부위(동일한 기관, 또는 상이한 기관내)에 존재한다. 특정 유형의 종양에서, 종양 이종성의 존재는 불량한 예후의 지시인자이다(참조: Yoon HH 등 J Clin Oncol 2012; 30: 3932-3938; Merlo LMF 등 Cancer Prev Res 2010; 3: 1388-1397). 특정 유형의 종양에서, 종양 이종성 정도가 높을 수록, 표적화된 처리 후 내성 클론의 발달 또는 종양 진행 기회가 더 높아질 수 있다.

[0286] 암이 하나의 종양 세포의 클론성 확장으로부터 발생하는 것으로 여겨진다고 해도, 암의 성장 및 발달은 암의 상이한 부분에서 새롭고 상이한 돌연변이의 축적을 초래할 수 있다. 예를 들면, 암 환자가 전이로 발달하는 경우, 원래의 기관내 위치하는 종양 및 전이성 종양은 다수의 돌연변이를 공유할 수 있다. 그러나, 2개 부위의 암 세포는 또한 다른 종양 부위에서는 부재하는 돌연변이의 유일한 세트를 수반할 수 있다. 2개의 부위에 의해 공유된 돌연변이는 하나의 종양 부위에서 유일하게 관찰되는 돌연변이보다 더 높은 농도로 존재하는 것으로 예측된다.

[0287] A. 실시예

[0288] 본 발명자는 쌍방 난소암 및 유방암을 가진 환자의 혈액 혈장을 분석하였다. 난소 종양 둘 다는 심각한 선암종이었다. 좌측 종양은 6cm로 측정되었고 우측 종양은 가장 긴 직경이 12cm로 측정되었다. 결장 및 그물막에서 다수의 전이성 병변이 존재하였다. 백혈구로부터 추출된 DNA를 업자(Illumina)로부터 합성에 의한 서열분석 플랫폼(sequencing-by-synthesis platform)을 사용하여 평균 44배의 반수체 게놈 포함으로 서열분석하였다. 단지 1개의 대립형질, 즉, 동형접합성을 나타내는 뉴클레오타이드 위치를 혈장 속에서 단일의 뉴클레오타이드 돌연변이에 대해 추가로 분석하였다.

[0289] DNA를 좌측 및 우측 종양의 4개의 상이한 부위로부터 추출하고 Illumina 서열분석 플랫폼을 사용하여 서열분석하였다. 2개 부위(부위 A 및 B)는 우측 종양으로부터 기원하였으며 다른 2개의 부위(부위 C 및 D)는 좌측 종양으로부터 기원하였다. 부위 A 및 B는 대략 4 cm 떨어져 있었다. 부위 C와 부위 D 사이의 거리 또한 대략 4 cm이었다. 혈장 시료를 난소 종양의 외과적 절개 전 및 후에 환자로 부터 수집하였다. 이후에, DNA를 환자의 혈장으로부터 추출하였다. 부위 A, B, C 및 D, 및 또한 혈장 시료로부터 종양의 서열분석 깊이는 표 9에 나타난다.

표 9

[0290] 부위 A, B, C 및 D로부터 종양의 서열분석 깊이

시료	원래의 서열분석 판독물의 수	정렬된 판독물의 수	반수체 게놈 포함의 배수
연막으로터의 구성적 DNA	1,091,250,072	876,269,922	43.81
우측 난소 종양(부위 A)	1,374,495,256	1,067,277,229	53.36
우측 난소 종양(부위 B)	934,518,588	803,007,464	40.15
좌측 난소 종양(부위 C)	1,313,051,122	1,036,643,946	51.83
좌측 난소 종양(부위 D)	1,159,091,833	974,823,207	48.74
앞서 수집된 혈장 시료	988,697,457	741,982,535	37.10
후에 수집된 혈장 시료	957,295,879	564,623,127	28.23

[0291] 현재의 실시예에서, 단일의 종양 관련된 단일의 뉴클레오타이드 돌연변이를 정의하기 위하여, 뉴클레오타이드 위치를 종양 조직에서 적어도 20회 및 구성적 DNA에서 30회 서열분석한다. 다른 구현예에서, 다른 서열분석 깊이, 예를 들면, 35, 40, 45, 50, 60, 70, 80, 90, 100 및 >100배를 사용할 수 있다. 서열분석 비용의 감소는 보다 더 용이하게 수행될 증가된 깊이를 허용할 수 있다. 뉴클레오타이드 위치는 구성적 DNA 내에서 동형접합성인 반면 뉴클레오타이드 변화는 종양 조직 내에서 관찰된다. 종양 조직 내 뉴클레오타이드 변화의 발생을 위한 기준은 종양 조직 내 특수한 뉴클레오타이드 위치의 전체 서열분석 깊이에 의존한다. 20 내지 30배로부터의 뉴클레오타이드 포함의 경우, 뉴클레오타이드 변화(컷오프 값)의 발생은 적어도 5배이다. 31 내지 50배로부터의 포함의 경우, 뉴클레오타이드 변화의 발생은 적어도 6배이다. 51 내지 70배의 포함의 경우, 발생 요구도는 적어도 7배이다. 이들 기준은 푸아송 분포를 사용하여 실제 돌연변이 및 거짓 양성 유전자좌의 예측된 수를 검출하는 민감성의 예측으로부터 기원한다.

[0292] 도 17은 발생의 상이한 요건 및 서열분석 깊이의 예측된 민감성을 나타내는 표 1700이다. 당해 민감성은 특수한 컷오프를 사용하는 특수한 배 깊이에서 검출된 실제 돌연변이의 수에 상응할 수 있다. 서열분석 깊이가 클수록, 보다 많은 돌연변이 서열 판독물이 수득될 것이므로, 제공된 컷오프에 대해 검출된 돌연변이일 확률이 더 크다. 보다 큰 컷오프 값의 경우, 당해 기준이 보다 더 엄격하므로, 돌연변이가 검출될 수 있는 확률은 거의 없다.

[0293] 도 18은 상이한 컷오프 및 상이한 서열분석 깊이에 대한 거짓 양성 유전자좌의 예측된 수를 나타내는 표 1800이다. 거짓 양성률의 수는, 보다 많은 서열 판독물이 수득되므로, 서열분석 깊이의 증가와 함께 증가한다. 그러나, 거짓 양성률은 5 이상의 컷오프, 심지어 70의 서열분석 깊이까지 예측되지 않는다. 다른 구현예에서, 발생의 상이한 기준을 사용하여 요구된 민감성 및 특이성을 달성할 수 있다.

[0294] 도 19는 상이한 종양 부위에서 검출된 돌연변이의 수를 나타내는 수형도이다. 돌연변이는 종양을 직접 서열분석하여 결정하였다. 비록 이들이 단지 4cm 떨어져 있다고 해도, 부위 A는 이러한 종양에 대해 특이적인 71개의 돌연변이를 가지고 있으며, 부위 B는 122개의 부위-특이적인 돌연변이를 가지고 있다. 부위 A 및 B 둘 다에서 10개의 돌연변이가 관찰되었다. 비록 이들이 단지 4cm 떨어져 있다고 해도, 부위 C는 이러한 종양에 대해 특이적인 168개의 돌연변이를 가지고 있으며, 부위 D는 248개의 부위-특이적인 돌연변이를 가지고 있다. 12개의 돌연변이가 부위 C 및 D 둘 다에서 관찰되었다. 상이한 종양 부위에 대해 돌연변이적 프로파일에 있어서 유의적인 이종성이 존재한다. 예를 들면, 248개의 돌연변이가 부위 D 종양에서 단지 검출되었지만 다른 3개의 종양 부위에서는 검출되지 않았다. 총 2,129개의 돌연변이가 모든 부위에 걸쳐 관찰되었다. 따라서, 많은 돌연변이가 사이한 종양 중에서 공유되었다. 따라서, 7개의 SNV 그룹이 존재하였다. 카피 수 비정상의 측면에서 이들 4개 영역 중에 관찰가능한 차이는 존재하지 않았다.

[0295] 도 20은 처리 전 및 처리 후 혈장 시료 속에서 종양 기원한 돌연변이를 수반하는 단편의 수를 나타내는 표 2000이다. 각각의 돌연변이를 수반하는 종양 유래 DNA의 추론된 분획 농도 또한 나타내었다. 돌연변이의 범주는, 돌연변이가 검출된 종양 부위(들)을 말한다. 예를 들어, 범주 A 돌연변이는 부위 A내에 단지 존재하는 돌연변이를 말하는 반면, 범주 ABCD 돌연변이는 모든 4개의 종양 부위 내에 존재하는 돌연변이를 말한다.

[0296] 모든 4개의 종양 부위에 존재한 2,129개 돌연변이의 경우, 2,105개(98.9%)가 적어도 하나의 혈장 DNA 단편에서 검출가능하였다. 한편, 4개의 종양 부위 중 단지 하나에 존재한 609개의 돌연변이의 경우, 단지 77개(12.6%)가 적어도 하나의 혈장 DNA 단편 속에서 검출가능하였다. 따라서, 혈장 내 단일 뉴클레오타이드 돌연변이의 정량

화는 종양 조직 내 이들 돌연변이의 상대적인 풍부성을 반영하기 위해 사용될 수 있다. 당해 정보는 암 이종성의 연구를 위해 유용할 수 있다. 당해 실시예에서, 잠재적인 돌연변이는, 이것이 서열분석 데이터 내에 1회 존재하였던 경우에만 요청된다.

[0297] 순환하는 종양 DNA의 분획 농도는 각각의 SNV 그룹으로 결정하였다. 모든 4개의 영역(즉, 그룹 ABCD)에 의해 공유된 SNV에 의해 측정된 것으로서, 수술 전 및수술 후 혈장 속에서 종양 DNA의 분획 농도는 각각 46% 및 0.18%였다. 이들 후자의 퍼센트는 GAAL 분석에서 수득된 것, 46% 및 0.66%와 잘 상호 관련되었다. 모든 4개의 영역(즉, 그룹 ABCD)에 의해 공유된 돌연변이는 혈장에 대한 종양 유래 DNA의 최대 분획 기여에 기여하였다.

[0298] 그룹 AB 및 CD로부터 SNV로 측정된 수술전 혈장 내 종양-기원한 DNA의 분획 농도는 각각 9.5% 및 1.1%였다. 이들 농도는 우측 및 좌측 난소 종양의 상대적 크기와 일치하였다. 영역-유일한 SNV로 측정된 종양 유래 DNA의 분획 농도(즉, 그룹 A, B, C 및 D내의 것들)는 일반적으로 낮았다. 이들 데이터는, 암 환자에서 전체 종양 부하의 정밀한 측정을 위해, 게놈전체 선건 시도(genomewide shotgun approach)의 사용이 특이적인 종양 관련된 돌연변이를 표적화하는 보다 전통적인 시도와 비교하여, 보다 대표적인 사진을 제공할 수 있음을 제안한다. 후자의 시도의 경우, 종양 세포의 서브세트만이 표적화된 돌연변이를 소유하는 경우, 표적화된 돌연변이를 소유하지 않는 종양 세포에 의해 유발된 임박한 재발 또는 질병 진행에 관한 중요한 정보를 손실할 수 있거나 처리-내성 클론의 출현을 손실할 수 있다.

[0299] 도 21은 단일의 종양 부위에서 검출된 돌연변이 및 모든 4개의 종양 부위에서 검출된 돌연변이에 대한 혈장 내 발생의 분포를 나타내는 그래프 (2100)이다. 막대 그래프 (2100)은 2개 유형의 돌연변이: (1) 단지 하나의 부위에서 검출된 돌연변이 및 (2) 모든 4개의 돌연변이 부위에서 검출된 돌연변이에 대한 데이터를 나타낸다. 수평 축은, 돌연변이가 혈장 속에서 검출된 횟수이다. 수직 축은 수평 축에서 특수한 값에 상응하는 돌연변이의 퍼센트를 나타낸다. 예를 들어, 유형 (1) 돌연변이중 약 88%는 혈장 속에서 단지 1회 나타났다. 알 수 있는 바와 같이, 1개 부위에서 나타난 돌연변이는 대부분 1회, 및 4회 이하로 검출되었다. 단일의 종양 부위 내에 존재하는 돌연변이는 모든 4개의 종양 부위에 존재하는 돌연변이와 비교하여 혈장속에서 매우 흔하지 않게 검출되었다.

[0300] 당해 기술의 한가지 적용은, 임상이가 상이한 부류의 돌연변이를 수반하는 종양 세포의 부하를 평가하도록 할 수 있다. 이들 돌연변이의 비율은 표적화된 제제로 잠재적으로 처리가능할 수 있다. 보다 높은 비율의 종양 세포에 의해 수반된 돌연변이를 표적화하는 제제는 보다 우세한 치료학적 효과를 가진 것으로 예측될 수 있다.

[0301] 도 22는 이형접합성 종양으로부터 기원한 돌연변이에 대한 혈장 내 예측된 발생 분포를 나타내는 그래프 (2200)이다. 당해 종양은 2개 그룹의 종양을 함유한다. 2개 부위가 각각의 난소 종양을 대표한다는 접근을 기초로 하여, 돌연변이의 1개 그룹은 모든 종양 세포 내에 존재하며 돌연변이의 다른 그룹은 종양 세포의 1/4로 단지 존재한다. 혈장 속에서 종양 유래 DNA의 전체 분획 농도는 40%인 것으로 추정된다. 혈장 시료는 뉴클레오타이드 위치당 50배의 평균 깊이로 서열분석되는 것으로 추정된다. 혈장 속에서 이러한 발생의 예측된 분포에 따라서, 모든 종양 조직 내에 존재하는 돌연변이는 혈장 속에서 이들의 발생에 의해 1/4 종양 세포 내에만 존재하는 돌연변이로부터 차별화될 수 있다. 예를 들어, 6회의 발생을 컷오프로 사용할 수 있다. 모든 종양 세포 내에 존재하는 돌연변이의 경우, 돌연변이의 92.3%는 혈장 속에 적어도 6회 존재할 수 있다. 대조적으로, 1/4 종양 세포에 존재하는 돌연변이의 경우, 돌연변이의 단지 12.4%만이 혈장 속에 적어도 6회 존재할 수 있다.

[0302] 도 23은 16명의 건강한 대조군 피검자에 대한 구현예의 특이성을 입증하는 표 (2300)이다. 이들의 혈장 DNA 시료는 30배의 중간 평균으로 서열분석되었다. 상기 난소 암 환자의 혈장 속에 존재한 돌연변이의 검출은 이들 건강한 피검자의 혈장 시료 속에서 수행하였다. 난소암 환자의 종양 속에 존재하는 돌연변이는 건강한 대조군 피검자의 혈장의 서열분석 데이터 내에서 매우 흔하게 검출되었으며 돌연변이 범위의 어느 것도 >1%의 명백한 분획 농도를 가지지 않았다. 이들 결과는, 당해 검출 방법이 매우 특이적임을 나타낸다.

[0303] B. 방법

[0304] 도 24는 본 발명의 구현예에 따라서 피검자의 하나 이상의 종양의 이형접합성을 분석하기 위한 방법 (2400)의 흐름도이다. 방법 (2400)의 특정 단계는 본원에 기술된 바와 같이 수행할 수 있다.

[0305] 블록 (2410)에서, 피검자의 구성적 게놈을 수득한다. 블록 (2420)에서, 하나 이상의 서열 태그를 피검자의 생물학적 시료내 다수의 DNA 단편 각각에 대해 접수하였으며, 여기서 생물학적 시료는 무세포 DNA를 포함한다. 블록 (2430)에서, 게놈 위치를 서열 태그에 대해 측정한다. 블록 (2440)에서,서열 태그를 구성적 게놈과 비교하여 제1 유전자좌의 첫 번째 수를 결정한다. 각각의 제1 유전자좌에서, 상기 구성적 게놈에 대해 서열 변이를

갖는 서열 태그의 수는 컷오프 값 초과이고, 여기서 컷오프 값은 1 이상이다.

[0306] 블록 (2450)에서, 하나 이상의 종양의 이중성의 척도를 제1 계층 위치의 세트의 각각의 첫 번째 수를 기초로 계산하였다. 하나의 국면에서, 당해 척도는 종양에 의해 공유되지 않은 돌연변이의 수에 대하여 종양에 의해 공유된 돌연변이의 수를 나타내는 값을 제공할 수 있다. 여기서, 다양한 종양이 대상내 상이한 종양과 함께 단일의 대상물로서 존재할 수 있으며, 이는 종양 내 이중성으로 일반적으로 불리는 것이 무엇인지를 나타낼 수 있다. 당해 척도는 또한, 일부 돌연변이가 많은 또는 대부분의 종양 내에 존재하는 돌연변이와 비교하여 하나 또는 수개의 종양 내에 존재하는지에 관한 것일 수 있다. 이중성의 하나 이상의 척도를 계산할 수 있다.

[0307] 블록 (2460)에서, 이중성 척도를 역치값과 비교하여 이중성의 수준의 분류를 측정할 수 있다. 하나 이상의 측정치를 다양한 방법으로 사용할 수 있다. 예를 들면, 하나 이상의 이중성 척도 측정을 사용하여 종양 진행의 기회를 예측할 수 있다. 일부 종양에서, 이중성이 커질수록 치료(예를 들면, 표적화된 치료) 후 진행 기회가 더 높아지고 내성 클론의 출현 기회가 더 높아진다.

[0308] C. 종양 이중성 척도

[0309] 이중성 척도의 하나의 예는 혈장 속의 돌연변이의 상이한 그룹의 '농도 밴드'의 수이다. 예를 들어, 환자내에 2개의 우세한 종양 클론이 존재하는 경우, 및 이들 클론이 상이한 농도로 존재하는 경우, 본 발명자는 혈장 속에서 농도가 상이한 2개의 상이한 돌연변이를 찾을 것으로 예측할 수 있다. 이들 상이한 값은 상이한 돌연변이 세트에 대한 분획 농도를 측정함으로써 계산할 수 있으며, 여기서 각각의 세트는 종양중의 하나에 상응한다.

[0310] 이들 농도 각각은 '농도 밴드' 또는 '농도 부류'로 불릴 수 있다. 환자가 보다 많은 클론을 갖는 경우, 보다 많은 농도 밴드/부류가 관찰될 것이다. 따라서, 밴드가 많을 수록, 이중성이 많다. 농도 밴드의 수는 다양한 돌연변이에 대해 분획 농도를 플롯팅함으로써 관찰할 수 있다. 히스토그램은 각종 농도에 대해 제조될 수 있으며, 여기서 상이한 피크는 상이한 종양(또는 하나의 종양의 상이한 클론)에 상응한다. 거대 피크는 모든 또는 일부 종양(또는 종양의 클론)에 의해 공유되는 돌연변이에 대한 것일 것이다. 이들 피크를 분석하여 보다 작은 어느 피크가 결합하여 보다 큰 피크를 결정하는지를 측정할 수 있다. 예를 들면, 도 10b 및 11에 대한 핏팅 과정과 유사한 핏팅 과정을 사용할 수 있다.

[0311] 하나의 시행에서, 히스토그램은 분획 농도인 x-축 및 유전자좌의 양(예를 들면, 수 또는 비)인 y-축을 사용한 플롯이다. 모든 또는 일부 종양에 의해 공유된 돌연변이는 보다 높은 분획 농도를 생성할 수 있다. 당해 피크 크기는 특수한 분획 농도를 생성하는 유전자좌의 양을 나타낼 수 있다. 저 농도 및 고 농도에서 피크의 상대적 크기는 종양(또는 종양의 클론)의 이중성 정도를 반영할 수 있다. 고 농도에서 보다 높은 피크는, 대부분의 돌연변이가 대부분의 또는 모든 종양(또는 종양의 클론)에 의해 공유되며 보다 낮은 정도의 종양 이중성을 나타냄을 반영한다. 저 농도에서 피크가 더 클 경우, 대부분의 돌연변이는 수개의 돌연변이(또는 수개의 종양의 클론)에 의해 공유된다. 이는 보다 높은 정도의 종양 이중성을 나타낼 수 있다.

[0312] 존재하는 피크가 많을 수록, 존재하는 부위 특이적인 돌연변이가 더 많아진다. 각각의 피크는 상이한 돌연변이 세트에 상응할 수 있으며, 여기서, 돌연변이 세트는 종양의 서브세트(예를 들면, 위에서 나열한 바와 같이, 단지 1개 또는 2개의 종양)로부터 기원한다. 도 19의 예를 들면, 4개의 부위-피크만이 최소의 농도(종양의 상대적인 크기에 의존)를 가지는 경향이 있으며, AB 부위 및 CD 부위에 대한 2개의 피크, 및 모든 부위에 의해 공유된 돌연변이에 대한 피크와 함께, 총 7개의 피크가 존재할 수 있다.

[0313] 피크의 위치는 또한 종양의 상대적인 크기를 제공할 수 있다. 보다 큰 종양이 보다 많은 종양 DNA를 시료, 예를 들면, 혈장 내로 방출할 수 있으므로, 보다 큰 농도는 보다 큰 종양과 상호 관련될 수 있다. 따라서, 상이한 부류의 돌연변이를 수반하는 종양 세포의 부하를 평가할 수 있다.

[0314] 이중성 척도의 다른 예는 비교적 큰 변이체 관독물(예를 들면, 9-13)을 갖는 돌연변이 부위의 비의 비와 비교하여 비교적 적은 변이체 관독물(예를 들면, 4, 5, 또는 6)이다. 도 22를 다시 참조하면, 부위 특이적인 돌연변이는 보다 적은 변이체 관독물(이는 또한 보다 적은 분획 농도를 생성한다)을 가졌음을 알 수 있다. 공유 돌연변이는 보다 많은 변이체 관독물(이는 또한 보다 큰 분획 농도를 생성한다)을 갖는다. 10(보다 큰 수)에서 제2 비로 나눈 6(보다 작은 수)에서의 제1 비의 비율은 이중성 척도를 전달한다. 당해 비율이 작을 경우, 부위 특이적인 돌연변이가 거의 존재하지 않으므로, 이중성의 수준은 낮다. 당해 비율이 큰 경우(또는 공지된 표본으로부터 계산된 값보다 적어도 더 큰 경우), 이중성의 수준은 더 크다.

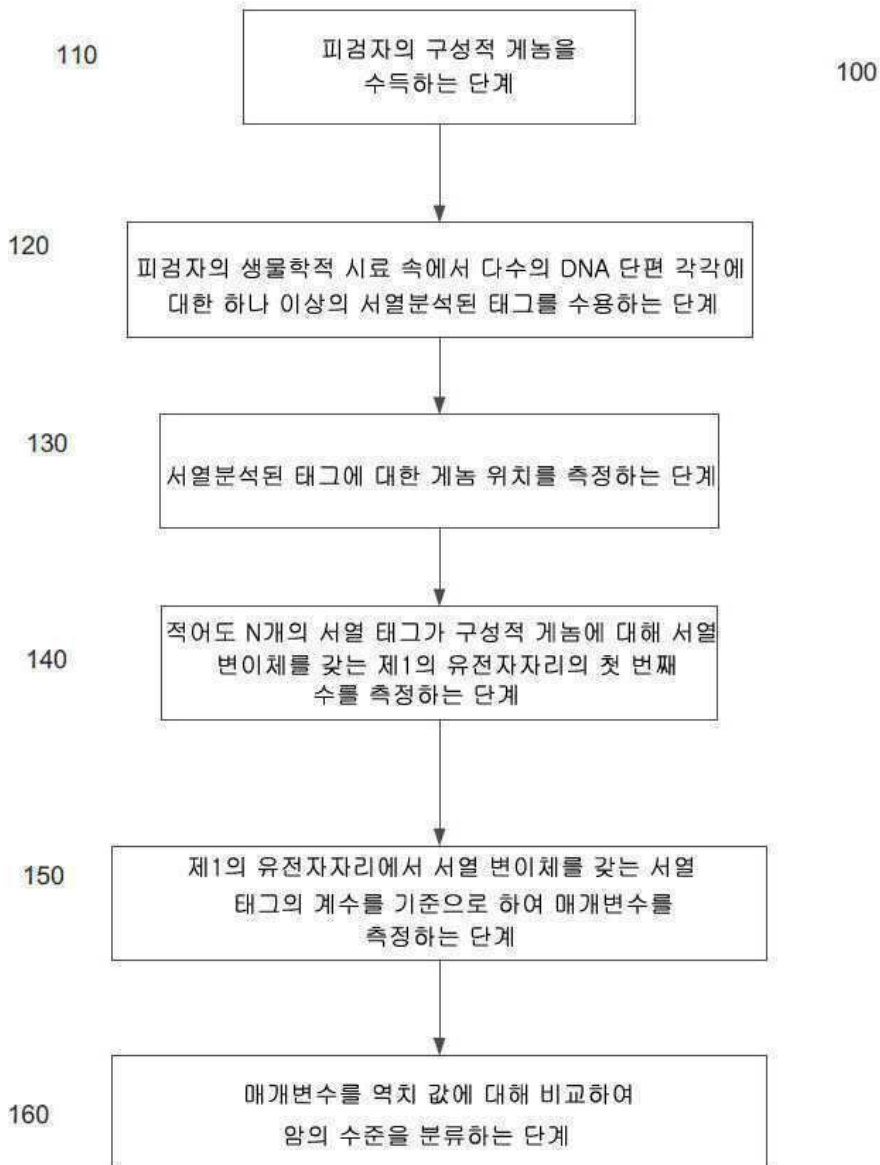
[0315] D. 역치의 측정

- [0316] 역치 값은, 이의 종양을 생검하여(예를 들면, 위에서 기술된 바와 같이) 이종성의 수준을 직접 결정하는 피검자로부터 측정될 수 있다. 당해 수준은 공유된 돌연변이에 대한 부위 특이적인 돌연변이의 비와 같은, 각종 방식으로 정의될 수 있다. 이후에, 생물학적 시료(예를 들면, 혈장 시료)를 분석하여 이종성 척도를 측정할 수 있으며, 여기서 생물학적 시료로부터의 이종성 척도는 종양의 세포를 직접 분석함으로써 측정된 이종성의 수준과 관련될 수 있다.
- [0317] 이러한 과정은 이종성 수준에 대한 역치의 보정을 제공할 수 있다. 시험 이종성 척도가 2개의 역치 사이에 속하는 경우, 이종성의 수준은 당해 역치에 상응하는 수준들 사이에 존재하는 것으로 평가될 수 있다.
- [0318] 하나의 구현예에서, 계산 곡선을 생검으로부터 측정된 이종성 수준과 혈장 시료(또는 다른 시료)로부터 측정된 상응하는 이종성 척도 사이에서 계산할 수 있다. 이러한 실시예에서, 이종성 수준은 수치적이며, 여기서 이들 수치 수준은 상이한 분류에 상응할 수 있다. 수치적 수준의 상이한 범위는 상이한 진단, 예를 들면, 암의 상이한 단계에 상응할 수 있다.
- [0319] *E. 게놈 표시로부터 분획 농도를 사용하는 방법*
- [0320] 종양 이종성을 또한 예를 들면, 방법 (1200)의 구현예를 사용하여 결정된 것으로서, 분획 농도를 사용하여 분석할 수 있다. 하나의 카피 손실을 나타내는 게놈 영역은 상이한 종양으로부터 올 수 있다. 따라서, 다양한 게놈 영역에 대해 측정된 분획 농도는, 증폭(또는 1개 카피 손실에 대한 결실)이 단지 하나의 종양 또는 다수의 종양에 존재하는지의 여부에 따라 상이할 수 있다. 따라서, 동일한 이종성 척도를 방법 (1200)의 구현예를 통해 측정된 분획 농도에 대해 사용할 수 있다.
- [0321] 예를 들면, 하나의 게놈 영역은 1개 카피 손실에 상응하는 것으로 확인될 수 있으며, 분획 농도는 이러한 게놈 영역에서 각각의 밀도(당해 각각의 밀도는 분획 농도로 사용될 수 있다)로부터 바로 측정할 수 있다. 히스토그램은, 밀도가 다양한 영역의 수를 계수함으로써 다양한 각각의 밀도로부터 측정할 수 있다. 단지 하나의 종양 또는 하나의 종양 클론 또는 하나의 종양 침착물이 특수 영역 내에서 획득을 갖는 경우, 이러한 영역의 밀도는 다수의 종양 또는 다수의 종양 클론 또는 다수의 종양 침착물(즉, 공유된 영역 내 종양 DNA의 분획 농도는 부위 특이적인 영역보다 더 클 수 있다)에서 획득을 가진 영역 내 밀도 미만일 수 있다. 따라서, 위에서 기술한 이종성 척도는 분획 농도의 분포를 나타낸 상이한 부위의 분획 농도와 같이, 다양한 영역 내 카피 수 획득 또는 손실을 사용하여 확인된 피크에 적용될 수 있다.
- [0322] 하나의 시행에서, 각각의 밀도가 히스토그램에 대해 사용된 경우, 분리된 획득 및 손실을 가질 수 있다. 획득을 나타내는 영역은 획득에 대해 바로 히스토그램을 생성함으로써 별도로 분석할 수 있으며, 별도의 히스토그램은 손실에 대해 바로 생성시킬 수 있다. 분획 농도를 사용하는 경우, 손실 및 획득의 피크를 함께 분석할 수 있다. 예를 들어, 분획 농도가 기준 밀도에 대한 차이(예를 들면, 절대 값으로서)를 사용하므로, 획득 및 손실에 대한 분획 농도는 동일한 피크에 기여할 수 있다.
- [0323] **XIII. 컴퓨터 시스템**
- [0324] 본원에 언급된 컴퓨터 시스템중 어느 것도 어떠한 적합한 수의 소시스템을 이용할 수 있다. 이러한 소시스템의 예는 컴퓨터 장치 (2500)에서 도 25에 나타낸다. 일부 구현예에서, 컴퓨터 시스템은 단일의 컴퓨터 장치를 포함하며, 여기서 소시스템은 컴퓨터 장치의 부품일 수 있다. 다른 구현예에서, 컴퓨터 시스템은 내부 부품과 함께, 각각 소 시스템인, 다수의 컴퓨터 장치를 포함할 수 있다.
- [0325] 도 25에 나타낸 소시스템은 시스템 버스(system bus) (2575)를 통해 서로연결되어 있다. 어댑터 2582를 나타내기 위해 커플링된, 프린터 (2574), 키보드 (2578), 고정 디스크 (2579), 모니터 (2576)과 같은 추가의 소시스템, 및 다른 것이 나타나 있다. 입력/출력(I/O) 컨트롤러 2571에 커플링되는, 주변 및 입력/출력(I/O) 장치는 일련 포트 (2577)과 같이, 당해 분야에 공지된 어떠한 다수의 수단에 의해 컴퓨터 시스템에 연결시킬 수 있다. 예를 들면, 일련의 포트 (2577) 또는 외부 인터페이스 (2581)(예를 들면, 이더넷, Wi-Fi 등)을 사용하여 컴퓨터 시스템 (2500)을 인터넷, 마우스 입력 장치, 또는 스캐너와 같은 광범위 네트워크에 연결시킬 수 있다. 시스템 버스 (2575)를 통한 상호연결은 중앙 프로세서 (2573)이 각각의 소시스템과 통신하도록 하고 시스템 기억장치 (2572) 또는 고정 디스크 (2579)로부터 지시의 실행, 및 또한 소시스템 사이의 정보의 교환을 조절할 수 있도록 한다. 시스템 기억장치 (2572) 및/또는 고정 디스크 (2579)는 컴퓨터 판독가능한 매체를 구현할 수 있다. 본원에 언급된 값들 중 어느 것도 하나의 부품으로부터 다른 부품으로 출력하는데 사용할 수 있으며 사용자에게 출력할 수 있다.

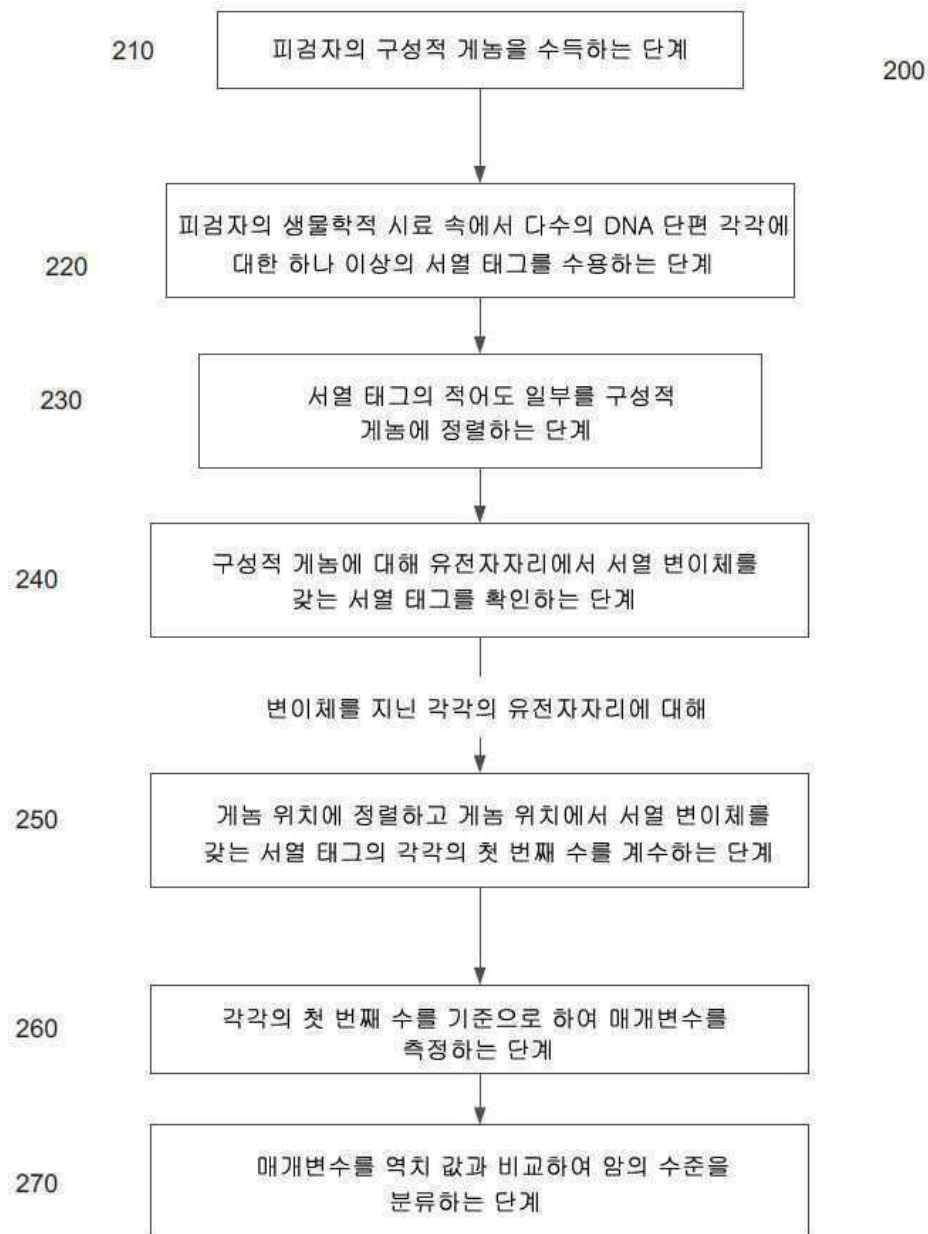
- [0326] 컴퓨터 시스템은 예를 들면, 외부 인터페이스 (2581)에 의해 또는 내부 인터페이스에 의해 함께 연결된, 다수의 동일한 부품 또는 소시스템을 포함할 수 있다. 일부 구현예에서, 컴퓨터 시스템, 소시스템, 또는 장치는 네트워크에 걸쳐 통신할 수 있다. 이러한 예에서, 하나의 컴퓨터는 고객으로 고려될 수 있고 다른 컴퓨터는 공급자로 고려될 수 있으며, 여기서 각각은 동일한 컴퓨터 시스템의 부분일 수 있다. 고객 및 공급자는 각각 다중 시스템, 소시스템, 또는 부품을 포함할 수 있다.
- [0327] 본 발명의 구현예 중 어느 것도 하드웨어(예를 들면, 적용 특이적인 집적회로 또는 필드 프로그램 가능한 게이트 어레이)의 형태로 및/또는 모듈러 또는 집적 방식으로 일반적으로 프로그램화가능한 프로세서가 장착된 컴퓨터 소프트웨어를 사용하여 제어 논리 형태로 삽입될 수 있다. 본원에 사용된 것으로서, 프로세서는 동일한 집적 칩 상에 다중-코어 프로세서, 또는 단일 회로판 또는 네트워크 상의 다중 프로세싱 장치를 포함한다. 본원에 제공된 기재내용 및 기술을 기본으로 하여, 당해 분야의 통상의 기술을 가진 자는 하드웨어 및 하드웨어와 소프트웨어의 조합을 사용하여 본 발명의 구현예를 시행하기 위한 다른 방식 및/또는 방법을 알고 인지할 것이다.
- [0328] 본 출원에 기술된 소프트웨어 부품 또는 기능 중 어느 것도 예를 들면, 통상의 또는 대상 기원한 기술을 사용하는, 예를 들면, 자바(Java), C++ 또는 Perl과 같은 어떠한 적합한 컴퓨터 언어를 사용하여 프로세서에 의해 실행될 소프트웨어 코드로서 시행될 수 있다. 소프트웨어 코드는 저장 및/또는 전송을 위한 컴퓨터 판독가능한 매체 상에 일련의 명령 또는 명령으로서 저장될 수 있으며, 적합한 매체는 임의접근 기억장치(RAM), 판독 전용 기억장치(ROM), 하드-드라이브 또는 플로피 디스크와 같은 자기 매체, 또는 콤팩트 디스크(CD) 또는 DVD(디지털 다기능 디스크), 플래시 메모리 등과 같은 자기 매체를 포함한다. 컴퓨터 판독가능한 매체는 이러한 저장 또는 전송 장치의 특정 조합일 수 있다.
- [0329] 이러한 프로그램은 또한 인터넷을 포함하는, 각종 프로토콜에 따른 유선, 광학 및/또는 무선 네트워크를 통한 전송을 위해 채택된 매개체 시그널을 사용하여 암호화되고 전송될 수 있다. 자체로서, 본 발명의 구현예에 따른 컴퓨터 판독가능한 매체는 이러한 프로그램으로 암호화된 데이터 시그널을 사용하여 생성될 수 있다. 프로그램 코드로 암호화된 컴퓨터 판독가능한 매체는 호환 장치와 함께 포장될 수 있거나 다른 장치(예를 들면, 인터넷 다운로드를 통해)로부터 별도로 제공될 수 있다. 어떠한 이런 컴퓨터 판독가능한 매체는 단일의 컴퓨터 프로그램 제품(예를 들면, 하드 드라이브, CD, 또는 전체 컴퓨터 시스템) 상에 또는 내에 속하며, 시스템 또는 네트워크에 상이한 컴퓨터 프로그램 제품에 또는 제품 속에 존재할 수 있다. 컴퓨터 시스템은 모니터, 프린터, 또는 본원에 언급된 결과 중 어느 것을 사용자에게 제공하기 위한 다른 적합한 디스플레이를 포함할 수 있다.
- [0330] 본원에 기술된 방법 중 어느 것도 하나 이상의 프로세서를 포함하는 컴퓨터 시스템을 사용하여 전체적으로 또는 부분적으로 수행할 수 있으며, 이는 단계를 수행하기 위해 배치될 수 있다. 따라서, 구현예는 잠재적으로 각각의 단계 또는 각각의 단계의 그룹을 수행하는 상이한 부품과 함께, 본원에 기술된 방법중 어느 것의 단계를 수행하기 위해 배치된 컴퓨터 시스템에 관한 것일 수 있다. 번호매긴 단계로서 나타낸다고 해도, 본원의 방법의 단계는 동시에 또는 상이한 순서로 수행될 수 있다. 또한, 이들 단계의 일부를 다른 방법으로부터의 다른 단계의 일부와 함께 사용할 수 있다. 또한, 단계 모두 또는 일부는 임의적일 수 있다. 또한, 방법들 중 어느 것의 단계들 중 어느 것도 이들 단계를 수행하기 위한 모듈, 회로, 또는 다른 수단으로 수행할 수 있다.
- [0331] 특수 구현예의 구체적인 세부사항은 본 발명의 구현예의 취지 및 영역으로부터 벗어남이 없이 어떠한 적합한 방식으로 합할 수 있다. 그러나, 본 발명의 다른 구현예는 각각의 개개 국면에 관한 구체적인 구현예, 또는 이들 개개 국면의 구체적인 조합에 관한 것일 수 있다.
- [0332] 본 발명의 예시적인 구현예의 상기 기술은 나열 및 기술의 목적으로 나타낸다. 본 발명을 기술된 정밀한 형태로 철저하게 하거나 제한하는 것을 의도하지 않으며, 많은 변형 및 변화가 상기 교시의 측면에서 가능할 수 있다. 구현예들은, 본 발명의 원리 및 이의 실제 적용을 가장 우수하게 설명함으로써 당해 분야의 숙련가가 본 발명을 다양한 구현예에서 및 고려된 특수 용도에 적합한 것으로서 다양한 변형으로 가장 우수하게 이용할 수 있도록 선택되어 기술되었다.
- [0333] 단수("a", "an" 또는 "the")의 인용은 반대로 구체적으로 나타내지 않은 한, "하나 이상"을 의미하는 것으로 의도된다.
- [0334] 본원에 언급된 모든 특허, 특허원, 공보, 및 설명은 모든 목적을 위해 이의 전문이 참조로 혼입된다. 어느 것도 선행 기술인 것으로 허용되지 않는다.

도면

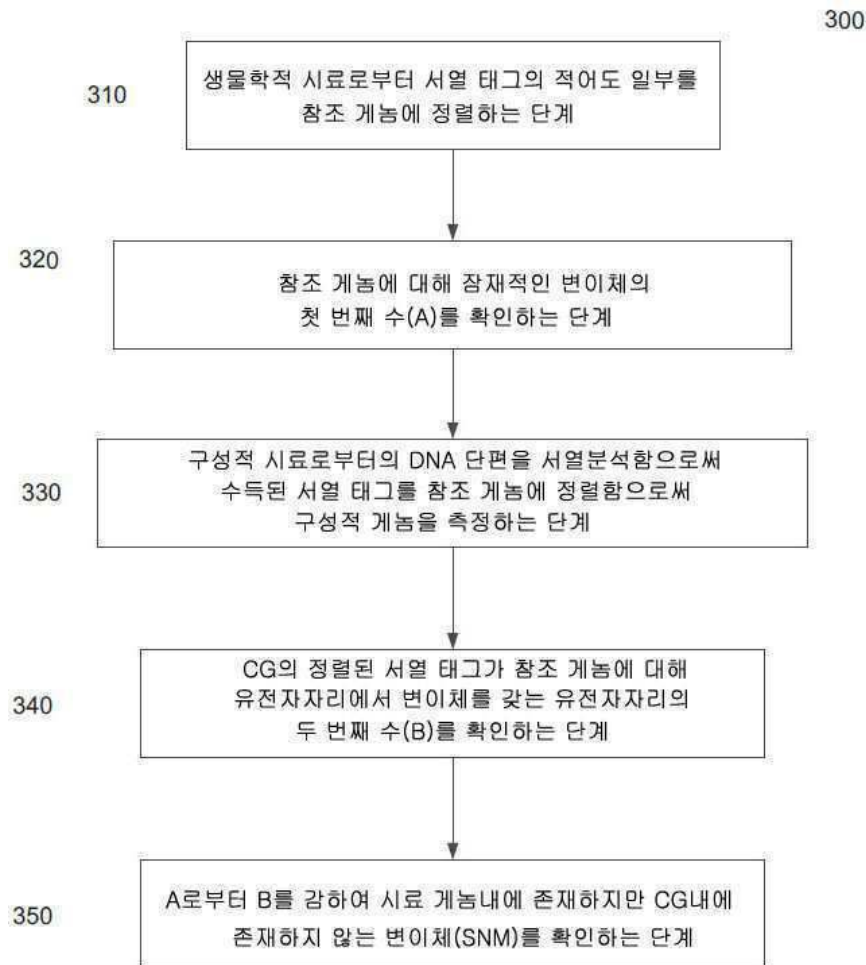
도면1



도면2



도면3



도면4

관측된 횟수	1x		2x		3x		4x		5x		6x	
	서열분석 오지에 대한 거짓 양성 유전자지리의 예측된 수	확인된 돌연변이의 예측된 수	서열분석 오지에 대한 거짓 양성 유전자지리의 예측된 수	확인된 돌연변이의 예측된 수	서열분석 오지에 대한 거짓 양성 유전자지리의 예측된 수	확인된 돌연변이의 예측된 수	서열분석 오지에 대한 거짓 양성 유전자지리의 예측된 수	확인된 돌연변이의 예측된 수	서열분석 오지에 대한 거짓 양성 유전자지리의 예측된 수	확인된 돌연변이의 예측된 수	서열분석 오지에 대한 거짓 양성 유전자지리의 예측된 수	확인된 돌연변이의 예측된 수
1	2,998,500	146	1,499	4	0	0	0	0	0	0	0	0
10	29,850,499	1,180	45,000	271	40	43	0	5	0	1	0	0
20	59,403,980	1,896	190,000	793	380	241	1	57	0	11	0	2
40	117,631,683	2,594	780,000	1,782	3,293	970	10	429	0	158	0	50
60	174,706,399	2,851	1,770,000	2,403	11,407	1,730	54	1,058	0	554	0	252
80	230,650,961	2,945	3,160,000	2,725	27,387	2,286	176	1,700	1	1,113	0	645
100	285,487,746	2,980	4,950,000	2,879	53,900	2,626	436	2,205	3	1,679	0	1,152
120	339,238,690	2,993	7,140,000	2,948	93,613	2,814	913	2,546	7	2,145	0	1,663
140	391,925,294	2,997	9,730,000	2,978	149,193	2,911	1,703	2,755	15	2,481	0	2,098
160	443,568,633	2,999	12,720,000	2,991	223,307	2,959	2,922	2,873	30	2,701	0	2,426
180	494,189,366	3,000	16,110,000	2,996	318,620	2,981	4,700	2,936	55	2,835	1	2,653
200	543,807,741	3,000	19,900,000	2,999	437,800	2,992	7,187	2,969	94	2,912	1	2,799

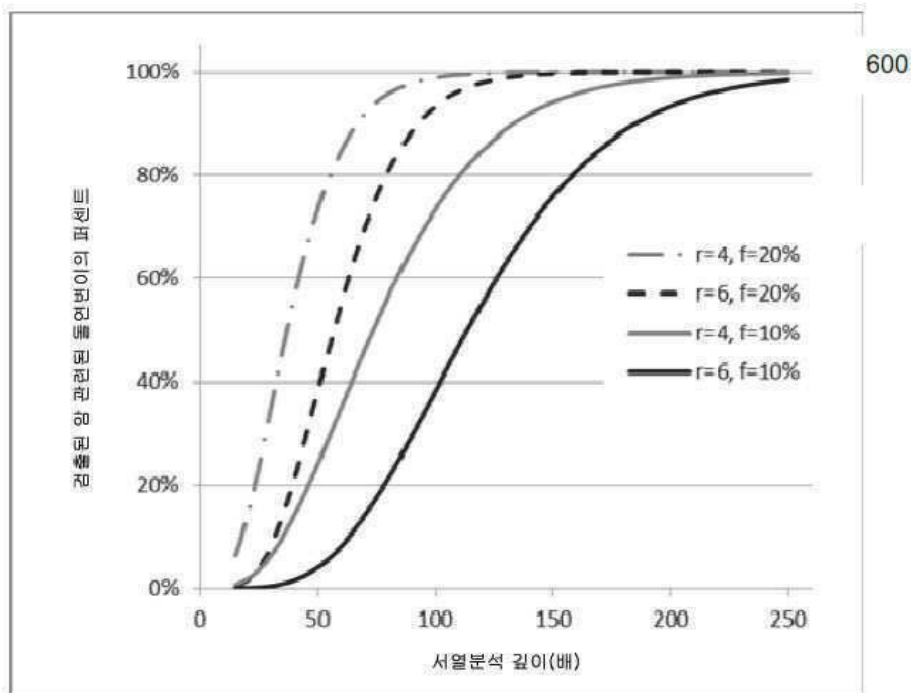
400

도면5

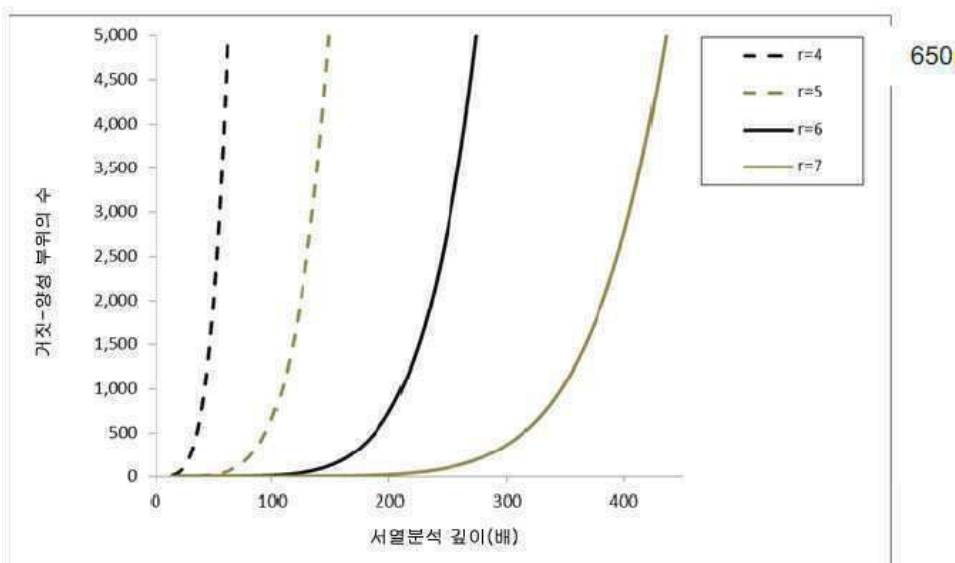
관찰된 횟수	1x		2x		3x		4x		5x		6x	
	서열분석 오지에 대한 거짓 양성 유전자지리의 예측된 수	확인된 돌연변이의 예측된 수	서열분석 오지에 대한 거짓 양성 유전자지리의 예측된 수	확인된 돌연변이의 예측된 수	서열분석 오지에 대한 거짓 양성 유전자지리의 예측된 수	확인된 돌연변이의 예측된 수	서열분석 오지에 대한 거짓 양성 유전자지리의 예측된 수	확인된 돌연변이의 예측된 수	서열분석 오지에 대한 거짓 양성 유전자지리의 예측된 수	확인된 돌연변이의 예측된 수	서열분석 오지에 대한 거짓 양성 유전자지리의 예측된 수	확인된 돌연변이의 예측된 수
서열분석 값이												
1	2,998,500	74	1,499	1	0	0	0	0	0	0	0	0
10	29,850,499	654	45,000	79	40	6	0	0	0	0	0	0
20	59,403,980	1,180	190,000	271	380	43	1	5	0	1	0	0
40	117,631,683	1,896	780,000	793	3,293	241	10	57	0	11	0	2
60	174,706,399	2,331	1,770,000	1,327	11,407	573	54	197	0	56	0	13
80	230,650,961	2,594	3,160,000	1,782	27,387	970	176	429	1	158	0	50
100	285,487,746	2,754	4,950,000	2,138	53,900	1,369	436	727	3	326	0	126
120	339,238,690	2,851	7,140,000	2,403	93,613	1,730	913	1,058	7	554	0	252
140	391,925,294	2,909	9,730,000	2,592	149,193	2,037	1,703	1,390	15	824	0	427
160	443,568,633	2,945	12,720,000	2,725	223,307	2,286	2,922	1,700	30	1,113	0	645
180	494,189,366	2,967	16,110,000	2,817	318,620	2,479	4,700	1,973	55	1,404	1	891
200	543,807,741	2,980	19,900,000	2,879	437,800	2,626	7,187	2,205	94	1,679	1	1,152

500

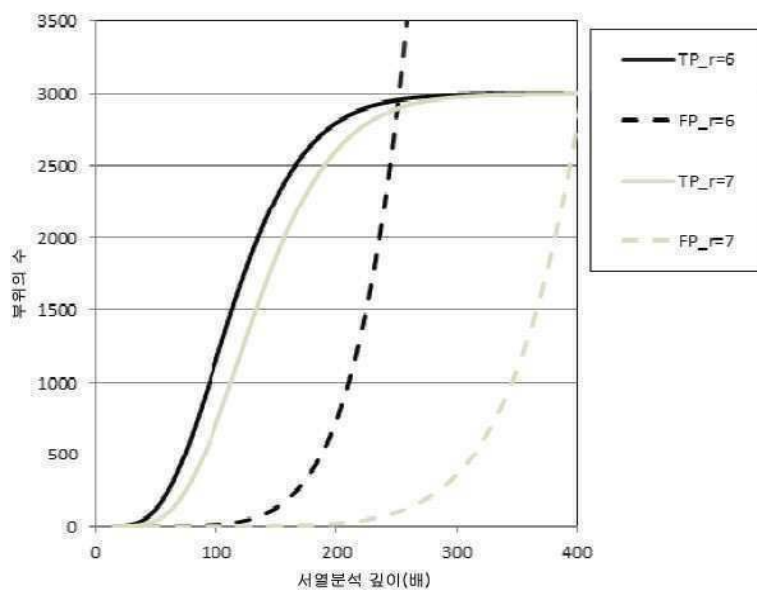
도면6a



도면6b

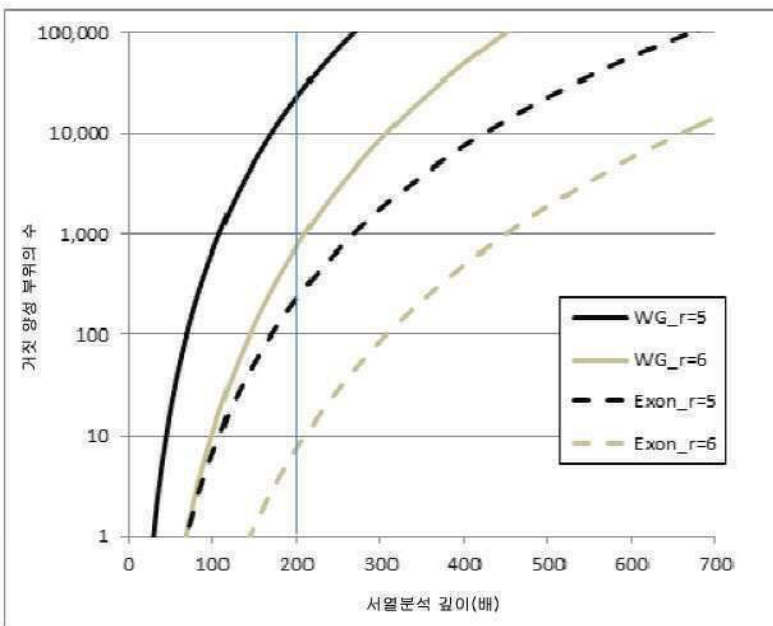


도면7a



700

도면7b



750

도면8

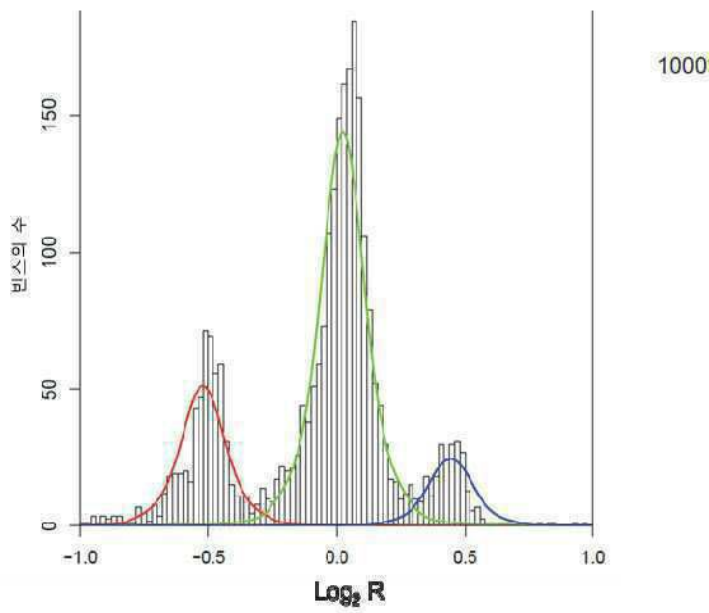
경우	종양 조직내에서 검출된 SNV의 수	시점	혈장으로부터 서열분석된 종양 관련된 SNV의 수(종양 조직에서 관찰된 SNV의 %)	SNV를 나타내는 혈장 DNA 서열 판독물의 수 (p)	야생형 서열을 나타내는 혈장 DNA 서열 판독물의 수 (q)	SNV 분석에 의한 종양 기원한 DNA의 유추된 분해 농도 $(\frac{2p}{p+q})$	GAAL 분석에 의한 종양 기원한 DNA의 유추된 분해 농도
HCC1	2,840	Tx 전	2,569 (94%)	11,389	31,602	53%	52%
		Tx 후	44 (1.5%)	91	46,898	0.4%	1.4%
HCC2	3,105	Tx 전	1,097 (35%)	1,490	57,865	5.0%	4.3%
		Tx 후	72 (2.3%)	206	66,692	0.6%	0.9%
HCC3	3,171	Tx 전	461 (15%)	525	48,886	2.1%	5.6%
		Tx 후	31 (1%)	67	58,862	0.2%	0.9%
HCC4	1,334	Tx 전	201 (15%)	248	18,527	2.6%	7.6%
		Tx 후	74 (5.5%)	149	22,144	1.3%	2.7%

도면9

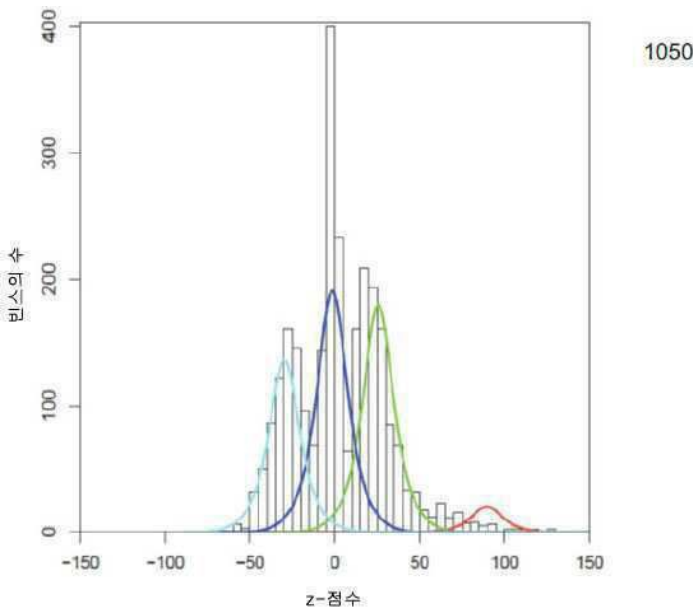
	HCC1 SNVs		HCC2 SNVs		HCC3 SNVs		HCC4 SNVs		모든 SNVs			
대조군	종양 관련인 SNV를 나타내는 혈장 DNA 서열 판독물	아생형 서열을 나타내는 혈장 DNA 서열 판독물	종양 관련인 SNV를 나타내는 혈장 DNA 서열 판독물	아생형 서열을 나타내는 혈장 DNA 서열 판독물	종양 관련인 SNV를 나타내는 혈장 DNA 서열 판독물	아생형 서열을 나타내는 혈장 DNA 서열 판독물	종양 관련인 SNV를 나타내는 혈장 DNA 서열 판독물	아생형 서열을 나타내는 혈장 DNA 서열 판독물	종양 관련인 SNV를 나타내는 혈장 DNA 서열 판독물	아생형 서열을 나타내는 혈장 DNA 서열 판독물	혈장 속의 모든 종양 관련된 SNV의 분획 농도 ($\frac{2p}{p+q}$)	
	C01	40	31,261	66	38,172	30	33,823	69	14,389	205	117,645	0.35%
	C02	60	34,368	112	42,660	37	37,462	58	16,097	267	130,587	0.41%
	C03	55	37,931	100	46,297	50	41,181	74	17,039	279	142,448	0.39%
	C04	41	40,277	87	48,441	39	43,541	87	17,358	254	149,617	0.34%
C05	42	31,655	71	38,787	46	34,566	70	14,821	229	119,829	0.38%	
C06	41	34,604	97	43,999	26	37,477	79	16,762	243	132,842	0.37%	
C07	52	38,664	118	47,200	61	41,840	72	17,601	303	145,305	0.42%	
C08	57	38,514	131	46,464	42	41,662	81	17,275	311	143,915	0.43%	
C09	41	34,945	115	43,760	33	38,039	81	16,503	270	133,247	0.40%	
C10	44	35,197	89	43,622	51	38,370	65	16,009	249	133,198	0.37%	
C11	41	36,693	95	45,108	39	39,732	72	16,801	247	138,334	0.36%	
C12	45	39,506	104	50,085	46	42,536	91	18,598	286	150,725	0.38%	
C13	36	33,799	71	40,309	39	36,241	70	14,450	216	124,799	0.35%	
C14	42	38,199	104	47,832	42	41,245	91	17,807	279	145,083	0.38%	
C15	49	36,106	90	43,129	41	38,432	73	16,154	253	133,821	0.38%	
C16	38	33,633	66	40,023	41	36,431	52	14,502	197	124,589	0.32%	
									평균		0.38%	

900

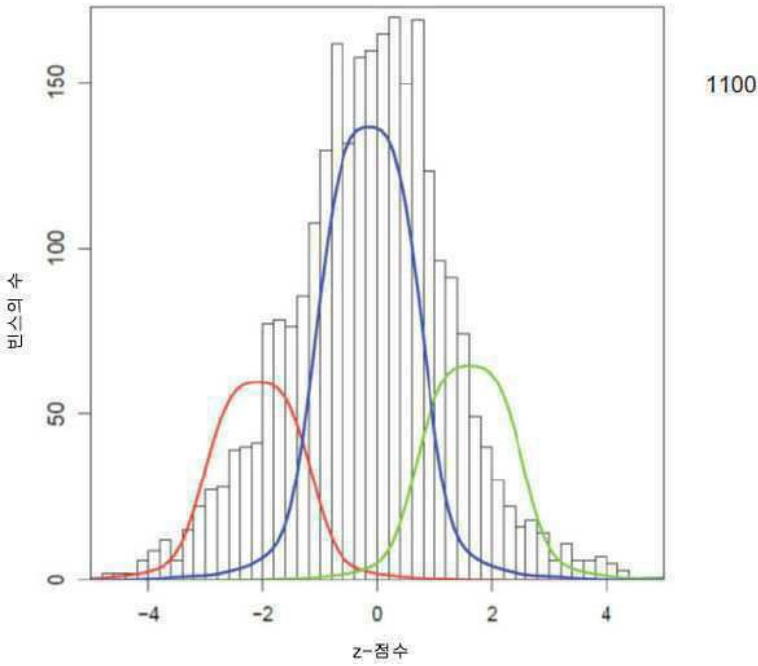
도면10a



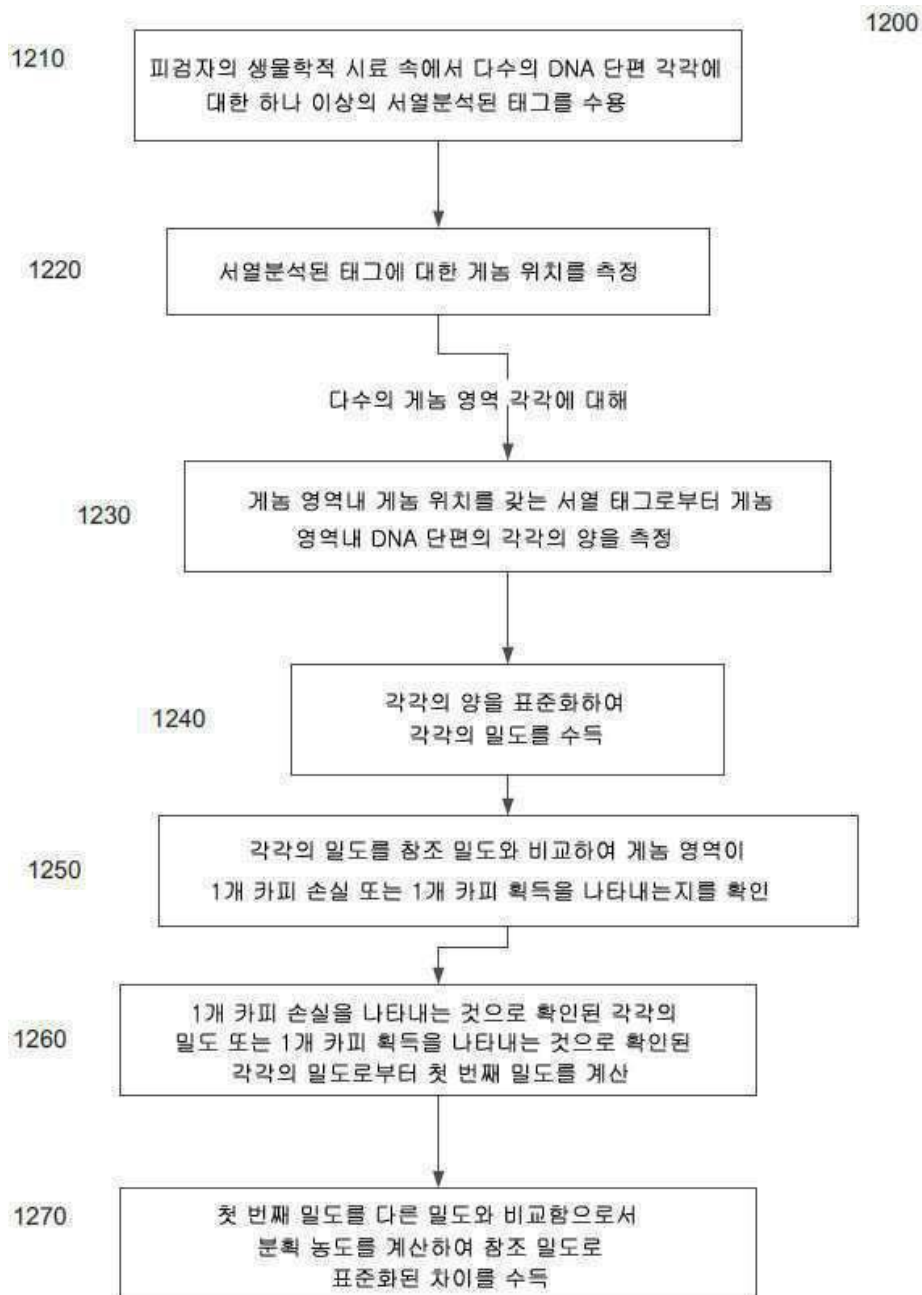
도면10b



도면11



도면12



도면13a

혈장 속에서 검출된 횟수 (N)	혈장 속에서 N회 검출된 잠재적 돌연변이의 수	혈장 속에서 ≥N 회 검출된 잠재적 돌연변이의 수	혈장 속에서 N회 검출된 중앙 돌연변이	혈장 속에서 ≥N회 검출된 중앙 돌연변이	N=1과 비교하여 검출된 돌연변이의 퍼센트	양성의 예측 값
1	3,605,085	3,623,970	61	2,064	100%	0.06%
2	13,518	18,885	35	2,003	97%	11%
3	2,029	5,367	71	1,968	95%	37%
4	766	3,338	83	1,897	92%	57%
5	437	2,572	140	1,814	88%	71%
6	313	2,135	161	1,674	81%	78%
7	265	1,822	165	1,513	73%	83%
8	261	1,557	192	1,348	65%	87%
9	231	1,296	185	1,156	56%	89%
10	191	1,065	160	971	47%	91%
11	179	874	165	811	39%	93%
12	148	695	137	646	31%	93%
13	122	547	109	509	25%	93%
14	99	425	92	400	19%	94%
15	74	326	71	308	15%	94%
16	54	252	49	237	11%	94%
17	39	198	39	188	9%	95%
18	35	159	33	149	7%	94%
19	24	124	23	116	6%	94%
20	20	100	20	93	5%	93%
21	18	80	17	73	4%	91%
22	11	62	10	56	3%	90%
23	9	51	9	46	2%	90%
24	8	42	5	37	2%	88%
25	6	34	6	32	2%	94%
26	7	28	5	26	1%	93%
27	8	21	8	21	1%	100%
28	1	13	1	13	1%	100%
29	4	12	4	12	1%	100%
30	8	8	8	8	0%	100%

1300

도면13b

혈장 속에서 검출된 횟수 (N)	혈장 속에서 N회 검출된 잠재적 돌연변이의 수	혈장 속에서 ≥N 회 검출된 잠재적 돌연변이의 수	혈장 속에서 N회 검출된 중양 돌연변이	혈장 속에서 ≥N회 검출된 중양 돌연변이	N=1과 비교하여 검출된 돌연변이의 퍼센트	양성의 예측 값
1	3,155,634	3,167,144	51	54	100%	<0.01%
2	9,835	11,510	2	3	6%	0.03%
3	1,123	1,675	1	1	2%	0.06%
4	314	552	-	-	-	-
5	132	238	-	-	-	-
6	49	106	-	-	-	-
7	22	57	-	-	-	-
8	11	35	-	-	-	-
9	6	24	-	-	-	-
10	5	18	-	-	-	-
11	4	13	-	-	-	-
12	2	9	-	-	-	-
13	2	7	-	-	-	-
14	3	5	-	-	-	-
15	-	2	-	-	-	-
16	-	2	-	-	-	-
17	1	2	-	-	-	-
18	-	1	-	-	-	-
19	1	1	-	-	-	-
20	-	-	-	-	-	-
21	-	-	-	-	-	-
22	-	-	-	-	-	-
23	-	-	-	-	-	-
24	-	-	-	-	-	-
25	-	-	-	-	-	-
26	-	-	-	-	-	-
27	-	-	-	-	-	-
28	-	-	-	-	-	-
29	-	-	-	-	-	-
30	-	-	-	-	-	-

1300

도면14a

서열분석 길이	처리 전 혈장			처리 후 혈장		
	규정된 서열분석 깊이를 지닌 뉴클레오타이드의 퍼센트	혈장 속에서 검출된 단일의 뉴클레오타이드 변이를 지닌 뉴클레오타이드 위치의 총 수	혈장에서 상응하는 변화를 가진 부위의 수	규정된 서열분석 깊이를 지닌 뉴클레오타이드의 퍼센트	혈장 속에서 검출된 단일의 뉴클레오타이드 변이를 지닌 뉴클레오타이드 위치의 총 수	혈장내 상응하는 변화를 지닌 부위의 수
<50	36.3%	9	8	30.5%	0	0
50 - 125	29.9%	15	15	31.2%	0	0
126 - 235	18.6%	21	21	20.8%	0	0
236 - 380	9.4%	7	6	10.9%	0	0
381 - 560	4.2%	2	2	4.8%	0	0
561 - 760	1.6%	3	3	1.7%	0	0
총		57	55		0	0

1400

도면14b

서열분석 길이	저리 진 환경			저리 후 환경		
	규정된 서열분석 깊이를 지닌 뉴클레오타이드의 퍼센트	환경 속에서 검출된 단일의 뉴클레오타이드 변이를 지닌 뉴클레오타이드 위치의 총 수	환경에서 상응하는 변화를 가진 부위의 수	규정된 서열분석 깊이를 지닌 뉴클레오타이드의 퍼센트	환경 속에서 검출된 단일의 뉴클레오타이드 변이를 지닌 뉴클레오타이드 위치의 총 수	환경내 상응하는 변화를 지닌 부위의 수
<50	30.0%	0	0	27.1%	0	0
50 - 125	33.6%	2	2	32.3%	0	0
126 - 235	21.6%	6	6	22.2%	0	0
236 - 380	10.2%	5	5	11.3%	0	0
381 - 560	3.7%	4	4	4.9%	0	0
561 - 760	0.9%	1	1	1.7%	0	0
총		18	18		0	0

1450

도면15a

서열분석 길이	처리 전 샘플			처리 후 샘플		
	규정된 서열분석 깊이를 지닌 뉴클레오타이드의 퍼센트	샘플 내에서 검출된 단일의 뉴클레오타이드 변이를 지닌 뉴클레오타이드 위치의 총 수	샘플에서 생성하는 변이를 가진 부위의 수	규정된 서열분석 깊이를 지닌 뉴클레오타이드의 퍼센트	샘플 내에서 검출된 단일의 뉴클레오타이드 변이를 지닌 뉴클레오타이드 위치의 총 수	샘플내 생성하는 변이를 지닌 부위의 수
<50	38.8%	0	0	18.5%	0	0
50 - 125	45.7%	0	0	28.6%	0	0
126 - 235	15.2%	0	0	25.1%	0	0
236 - 380	0.2%	0	0	16.2%	0	0
381 - 560	<0.1%	0	0	8.3%	0	0
561 - 760	<0.1%	0	0	3.4%	0	0
총		0	0		0	0

1500

도면15b

서열분석 길이	처리 전 월장			처리 후 월장		
	규정된 서열분석 깊이를 지닌 뉴클레오타이드의 퍼센트	월장 속에서 검출된 단일의 뉴클레오타이드 변이를 지닌 뉴클레오타이드 위치의 총 수	중앙에서 생용하는 변화를 가진 부위의 수	규정된 서열분석 깊이를 지닌 뉴클레오타이드의 퍼센트	월장 속에서 검출된 단일의 뉴클레오타이드 변이를 지닌 뉴클레오타이드 위치의 총 수	중앙내 생용하는 변화를 지닌 부위의 수
<50	26.7%	0	0	24.4%	0	0
50 - 125	35.9%	1	1	31.2%	1	1
126 - 235	24.7%	2	2	23.1%	0	0
236 - 380	10.4%	0	0	13.1%	0	0
381 - 560	2.2%	0	0	6.0%	0	0
561 - 760	0.1%	0	0	2.2%	0	0
총		3	3		1	1
						1550

도면16

서열분석 깊이	처리 전 혈장			처리 후 혈장		
	규정된 서열분석 깊이를 지닌 뉴클레오타이드의 퍼센트	혈장 속에서 검출된 단일의 뉴클레오타이드 변이를 지닌 뉴클레오타이드 위치의 총 수	종양에서 상응하는 변화를 가진 부위의 수	규정된 서열분석 깊이를 지닌 뉴클레오타이드의 퍼센트	혈장 속에서 검출된 단일의 뉴클레오타이드 변이를 지닌 뉴클레오타이드 위치의 총 수	종양내 상응하는 변화를 지닌 부위의 수
<50	31.1%	7	7	30.8%	0	0
50 - 125	30.3%	14	14	33.2%	1	0
126 - 235	20.5%	19	19	21.6%	0	0
236 - 380	11.3%	13	13	10.3%	0	0
381 - 560	5.2%	7	4	3.5%	0	0
561 - 760	1.8%	4	2	0.7%	0	0
총		64	59		1	0

1600

도면17

서열분석 깊이(배)	적어도 n회 동안 중앙 조직 서열분석 데이터에서 관찰되는 것으로서 돌연변이가 정의되는 경우의 민감성, 여기서, n=					
	4	5	6	7	8	9
20	73%	56%	38%	24%	13%	7%
21	77%	60%	43%	28%	16%	9%
22	80%	64%	47%	31%	19%	11%
23	83%	68%	51%	35%	22%	13%
24	85%	71%	55%	39%	26%	15%
25	87%	75%	59%	43%	29%	18%
26	89%	78%	63%	47%	33%	21%
27	90%	80%	67%	51%	36%	24%
28	92%	83%	70%	55%	40%	27%
29	93%	85%	73%	59%	44%	30%
30	94%	87%	76%	62%	48%	34%
31	95%	89%	78%	66%	51%	37%
32	96%	90%	81%	69%	55%	41%
33	96%	91%	83%	72%	58%	44%
34	97%	93%	85%	74%	61%	48%
35	97%	94%	87%	77%	65%	51%
36	98%	95%	88%	79%	68%	54%
37	98%	95%	90%	82%	70%	58%
38	99%	96%	91%	84%	73%	61%
39	99%	97%	92%	85%	76%	64%
40	99%	97%	93%	87%	78%	67%
41	99%	98%	94%	88%	80%	69%
42	99%	98%	95%	90%	82%	72%
43	99%	98%	96%	91%	84%	75%
44	100%	98%	96%	92%	86%	77%
45	100%	99%	97%	93%	87%	79%
46	100%	99%	97%	94%	89%	81%
47	100%	99%	98%	95%	90%	83%
48	100%	99%	98%	95%	91%	84%
49	100%	99%	98%	96%	92%	86%
50	100%	99%	99%	97%	93%	88%
51	100%	100%	99%	97%	94%	89%
52	100%	100%	99%	97%	95%	90%
53	100%	100%	99%	98%	95%	91%
54	100%	100%	99%	98%	96%	92%
55	100%	100%	99%	98%	96%	93%
56	100%	100%	99%	99%	97%	94%
57	100%	100%	100%	99%	97%	95%
58	100%	100%	100%	99%	98%	95%
59	100%	100%	100%	99%	98%	96%
60	100%	100%	100%	99%	98%	96%
61	100%	100%	100%	99%	98%	97%
62	100%	100%	100%	99%	99%	97%
63	100%	100%	100%	100%	99%	97%
64	100%	100%	100%	100%	99%	98%
65	100%	100%	100%	100%	99%	98%
66	100%	100%	100%	100%	99%	98%
67	100%	100%	100%	100%	99%	99%
68	100%	100%	100%	100%	99%	99%
69	100%	100%	100%	100%	100%	99%
70	100%	100%	100%	100%	100%	99%

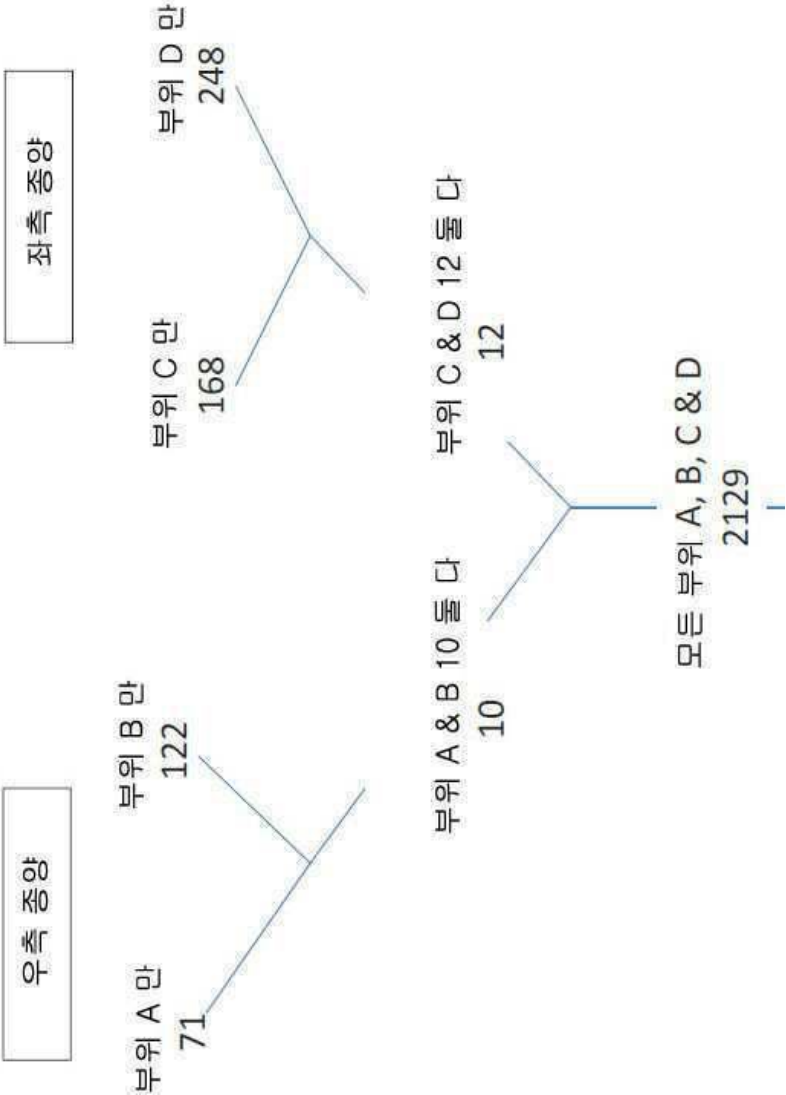
1700

도면18

서열분석 깊이(배)	적어도 n회 동안 중앙 서열분석 데이터에서 관찰되는 것으로서 돌연변이가 정의되는 경우의 전체 계층에서의 거짓 양성 수, 여기서 n=					
	4	5	6	7	8	9
20	1	0	0	0	0	0
21	1	0	0	0	0	0
22	1	0	0	0	0	0
23	1	0	0	0	0	0
24	1	0	0	0	0	0
25	1	0	0	0	0	0
26	2	0	0	0	0	0
27	2	0	0	0	0	0
28	2	0	0	0	0	0
29	3	0	0	0	0	0
30	3	0	0	0	0	0
31	3	0	0	0	0	0
32	4	0	0	0	0	0
33	5	0	0	0	0	0
34	5	0	0	0	0	0
35	6	0	0	0	0	0
36	7	0	0	0	0	0
37	7	0	0	0	0	0
38	8	0	0	0	0	0
39	9	0	0	0	0	0
40	10	0	0	0	0	0
41	11	0	0	0	0	0
42	12	0	0	0	0	0
43	14	0	0	0	0	0
44	15	0	0	0	0	0
45	17	0	0	0	0	0
46	18	0	0	0	0	0
47	20	0	0	0	0	0
48	22	0	0	0	0	0
49	24	0	0	0	0	0
50	26	0	0	0	0	0
51	28	0	0	0	0	0
52	30	0	0	0	0	0
53	33	0	0	0	0	0
54	35	0	0	0	0	0
55	38	0	0	0	0	0
56	41	0	0	0	0	0
57	44	0	0	0	0	0
58	47	0	0	0	0	0
59	51	0	0	0	0	0
60	54	0	0	0	0	0
61	58	0	0	0	0	0
62	62	0	0	0	0	0
63	66	0	0	0	0	0
64	71	0	0	0	0	0
65	75	0	0	0	0	0
66	80	0	0	0	0	0
67	85	0	0	0	0	0
68	90	0	0	0	0	0
69	96	0	0	0	0	0
70	102	0	0	0	0	0

1800

도면19

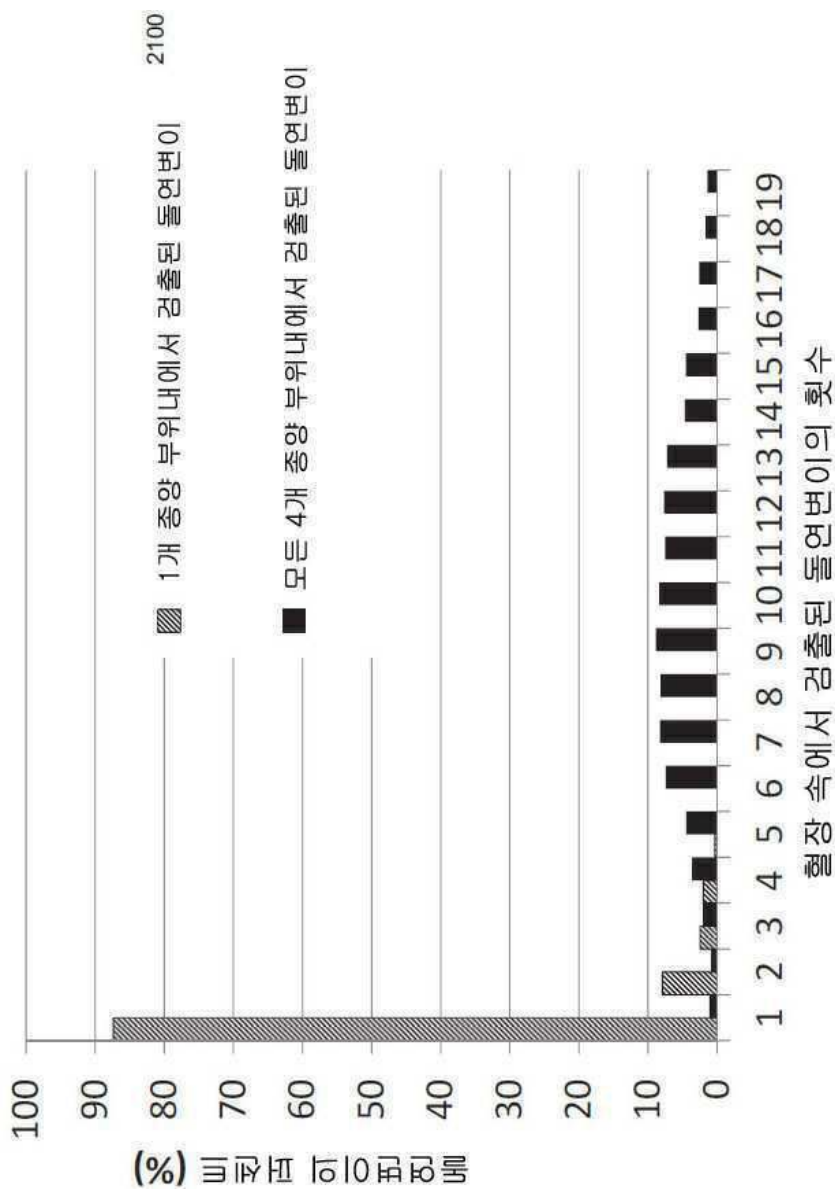


도면20

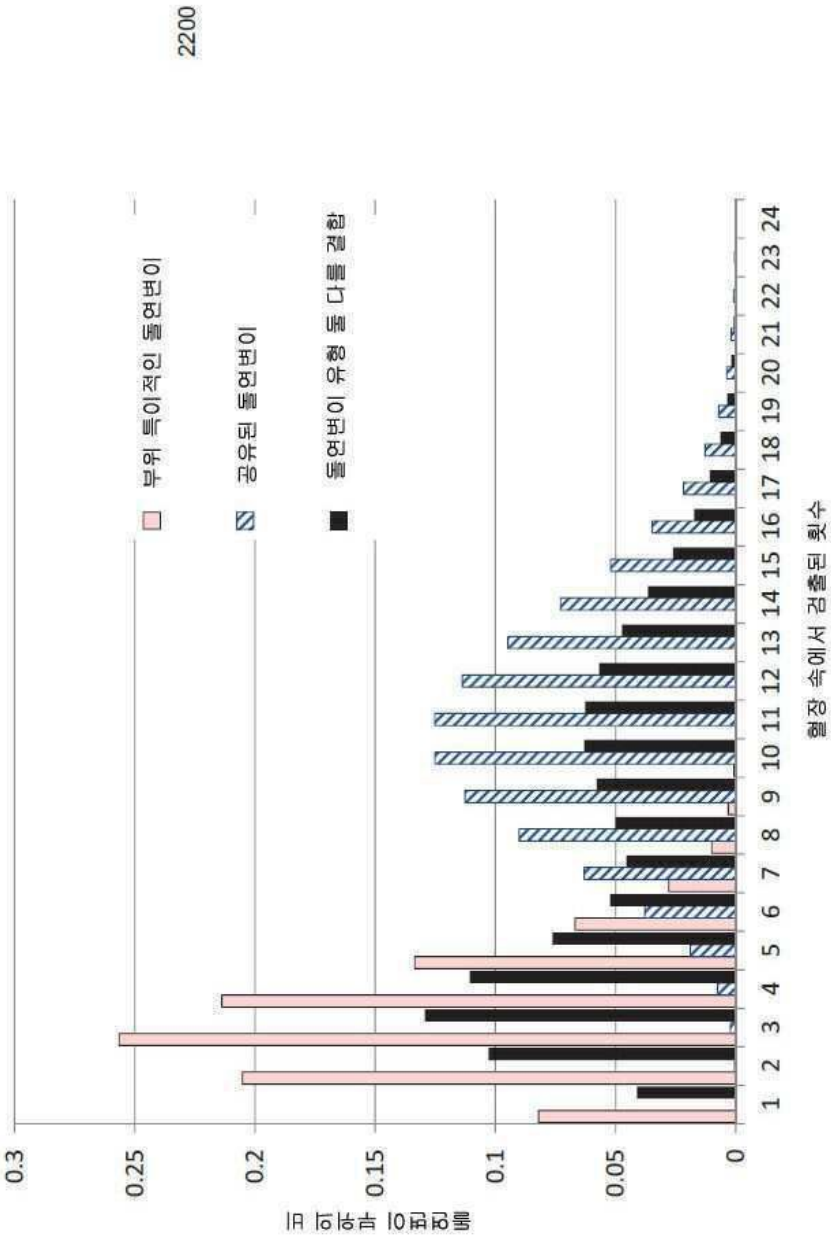
범주	돌연변이를 지닌 유전자자리 의 수	처리 전 혈장			처리 후 혈장		
		아생형 대립형질을 수반한 DNA 단편의 수	돌연변이체 대립형질을 수반한 DNA 단편의 수	중양 기원한 DNA의 분해 농도	아생형 대립형질을 수반한 DNA 단편의 수	돌연변이 대립형질을 수반한 DNA 단편의 수	중양 기원한 DNA의 분해 농도
A	71	3,321	37	2.20%	2,149	5	0.46%
B	122	5,633	8	0.28%	3,516	1	0.06%
C	168	8,507	51	1.19%	5,438	2	0.07%
D	248	10,880	26	0.48%	7,060	2	0.06%
AB	10	423	21	9.46%	297	0	0.00%
CD	12	569	3	1.05%	333	0	0.00%
ABCD	2129	70,940	21,344	46.26%	61,417	54	0.18%

2000

도면21



도면22

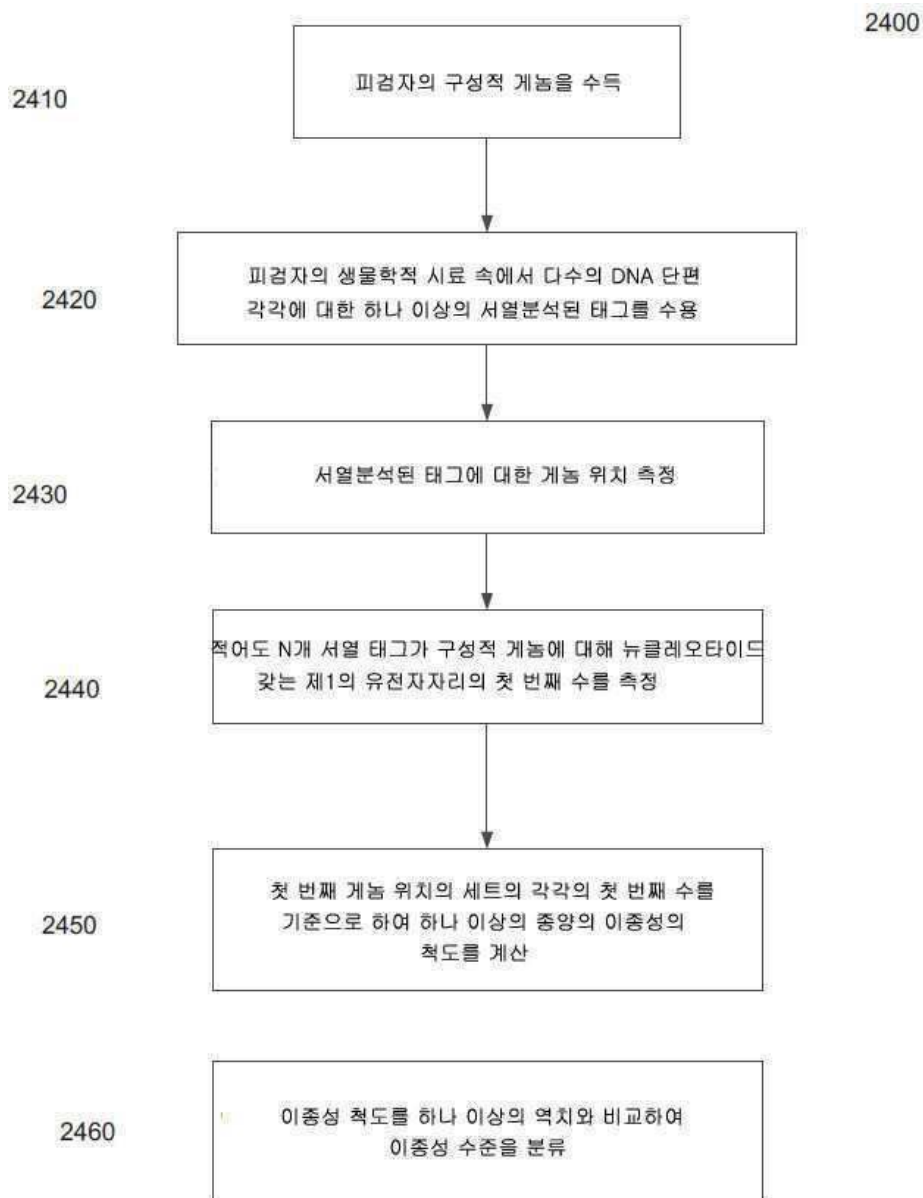


도면23

목적	건강한 대조군 피경자															
	유전자 자리의 수		TC1	TC2	TC3	TC4	TC5	TC6	TC7	TC8	TC9	TC10	TC11	TC12	TC13	TC14
			혈장 속의 돌연변이체 대립형질의 수(분획 농도)													
A	71	0	1 (0.17%)	1 (0.17%)	1 (0.18%)	0	1 (0.17%)	1 (0.18%)	1 (0.07%)	1 (0.20%)	0	6 (0.17%)	1 (0.10%)	0	1 (0.17%)	2 (0.38%)
B	122	0	0	0	0	0	0	0	1 (0.04%)	0	1 (0.11%)	3 (0.05%)	1 (0.10%)	1 (0.12%)	0	0
C	168	0	0	0	0	0	0	0	2 (0.06%)	0	2 (0.16%)	0	0	0	1 (0.07%)	0
D	248	0	0	0	1 (0.06%)	0	0	0	0	0	0	1 (0.01%)	0	0	1 (0.05%)	0
AB	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CD	12	0	0	0	0	0	0	0	0	0	0	1 (0.16%)	0	0	1 (0.92%)	0
ABCD	2129	11 (0.07%)	3 (0.02%)	2 (0.01%)	8 (0.06%)	5 (0.03%)	5 (0.03%)	5 (0.03%)	10 (0.02%)	4 (0.03%)	4 (0.03%)	19 (0.02%)	3 (0.02%)	0	1 (0.01%)	5 (0.03%)

2300

도면24



도면25

